Remarks

The Examiner has taken the position that the response filed April 2, 2003 was not fully responsive to the Office Action mailed October 2, 2002. In particular the Examiner states that:

> "[...] not all of the rejections of record were responded to, the double patenting rejection, for example. The applicant indicated that cancellation of all pending claims and filing new claims renders moot the outstanding objections and rejections. This is not correct because the new claims are drawn to the same or essentially the same subject matter that was rejected under various grounds in the prior Office Action and thus the new claims are subject to the same rejections as set forth in the prior Office Action. Accordingly, for the instant reply to be fully responsive, all rejections and objections of the prior Office Action must be responded to for the reply to be fully responsive."

Double patenting rejection:

Applicant acknowledges that the double patenting rejection was not addressed in the response filed April 2, 2003 and apologizes for this oversight. Claims 40, 45, 50, 57-59, 61, 63, 66, 68-70, 72-78, 80, 83-91 and 93-98 were rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over the claims of U.S. Patent No. 6,326,166 (the '166 patent). The '166 patent is commonly owned with the present application, and a Terminal Disclaimer obviating this rejection is enclosed herewith.

Response to claim objections and rejections as applied to new claims 99-119:

Applicant respectfully disagrees with the examiner's characterization of the response that was filed April 2, 2003. In particular, Applicant notes that the response included an extensive response to various claim objections and rejections *as they would apply to the newly added claims* (see pages 5-8). Applicant even explicitly stated:

> "The cancellation of all previously-pending claims renders moot the outstanding objections and rejections in this case. Nonetheless, Applicant has evaluated all of the arguments and art presented by the Examiner, and offers the following remarks evidencing the patentability of the present claims in light of these arguments and art."

More specifically, the rejections under 35 U.S.C. § 102 over Barbas et al. and Desjarlais et al. (see pages 4-6 of Office Action) were discussed on pages 5-6 of the response. The rejection under 35 U.S.C. § 103 over Ladner et al. in view of Park et al., Mitchell et al., Harrison and Schultz (see pages 15-21 of Office Action) was discussed on pages 6-8. Applicant acknowledges that the rejections under 35 U.S.C. § 103 over Park et al. in view of Mitchell et al., Harrison and Schultz (with or without Gossen et al., see pages 7-15 and 21-24 of Office Action) were not addressed in the response. These are therefore addressed in the following.

Rejections under 35 U.S.C. §103 over Park et al. in view of Mitchell et al., Harrison and Schultz:

Claims 40-70, 72, 89-92, 94-95 and 97 were rejected under 35 U.S.C. §103 as being unpatentable over Park et al. (*Proc. Natl. Acad. Sci. USA* 89:9094, 1992) in view of Mitchell et al. (*Science* 245:371, 1989), Harrison (*Nature* 353:715, 1991) and Schultz (*Nature* 240:426, 1988). Claims 40-70, 72, 89-92, 94-95 and 97 have been canceled. The cancelled claims have been replaced with new claims 99-119. Applicant respectfully traverses this rejection as applied to new claims 99-119.

New claims 99-119 are directed to a nucleic acid encoding a chimeric transcription factor (or a transcriptional regulatory protein) that has a chimeric nucleic acid binding domain and a transcriptional regulatory domain. The chimeric nucleic acid binding domain includes at least two nucleic acid binding motifs, at least one of which is a zinc finger, and the transcriptional regulatory protein has a different binding specificity than does a protein having only one of the motifs.

As acknowledged by the Examiner, none of the cited references teaches a nucleic acid that falls within the scope of claims 99-119. Thus, in order for the cited references to be properly relied upon in an obviousness rejection, they must include a suggestion or motivation to prepare the claimed nucleic acids; a reasonable expectation of success that the claimed nucleic acids will be made by those pursuing the suggestion or motivation; and a teaching or suggestion of every limitation of the claimed nucleic acids. Applicant respectfully submits that the cited references, taken alone or together with one another and/or with the general knowledge in the art at the time the present application was filed, fail to provide neither sufficient motivation for preparing the

claimed nucleic acids nor the requisite reasonable expectation of success.

Park et al. is relied upon as a primary reference. In particular, the Examiner cites Park et al. as teaching that at the time of the claimed invention it was "[...] within the ordinary skill in the art to stitch the DNA binding domains together from *any* proteins that recognize a specific DNA sequence by binding along the major groove, to recognize a composite binding site [...]" (Office Action, page 9). With regard to the use of DNA binding domains from different families and specific types of domains such as zinc-finger domains, the Examiner cites Park et al. as teaching that "[...] *any* combination of domains can be used, which would include heterologous ones [...]" (Office Action, page 11). Applicant respectfully submits that these characterizations dramatically overstate the teachings of Park et al.

Park et al. chemically synthesized a short peptide (31 amino acids) that corresponds to the basic region of v-Jun. The basic region of v-Jun was selected from within 14 leucine-zipper proteins, other DNA binding domains were not even considered in the selection process. A cysteine-containing tag was then introduced at the end of the peptide and two copies of the tagged peptide were covalently attached by chemical cross-linking. Thus, Park et al. showed that *a single* basic region from *a single* DNA-binding protein could be chemically synthesized and cross-linked to itself. The cross-linked entity bound DNA. Park et al. did not demonstrate the use of *any* nucleic acid-binding domain other than the particular basic region employed, and certainly did not demonstrate the "stitching" of heterologous binding domains nor the use of zinc-finger domains. If anything Park et al. teaches away from the use of zinc-finger domains by specifying the exclusion of DNA binding domains that include cysteine residues.

Furthermore, in order to place the teachings of Park et al. in their proper context one must note that two years before Park et al. published their work, Peter Kim and co-workers at MIT had already demonstrated the same methodology using the basic region from GCN4, a related leucine zipper protein that also homodimerizes and binds the exact same site as v-Jun (Talanian et al., *Science* 249:769, 1990, a copy of which is attached as **Exhibit A**). Park et al. acknowledge this and characterize their own work as "build[ing] upon the results of Kim and co-workers." Park et al. extended the work of Kim by selecting a different example of the *same class* of DNA binding proteins. Park et al. considered Kim's work to be limited to leucine zipper proteins. Other

researchers, reading Kim's work, came to the same conclusion (see, for example, Deng et al., *Proc. Natl. Acad. Sci. USA* 89:8572, 1992, a copy of which is attached as **Exhibit B**). One of ordinary skill in the art would consider Park et al.'s work to be similarly limited.

Applicant also notes that Park et al. does not demonstrate the construction of a *protein*, as Park et al. goes to some effort to chemically cross-link two peptides (as did Kim and co-workers), rather than to synthesize them together as a single polypeptide. Nor, obviously, does Park et al. describe a *nucleic acid* encoding a protein. The Examiner has asserted that the teachings of Park et al. are not limited to cross-linked peptides by pointing to a sentence that reads "use of the Gly-Gly-Cys linker is not essential in the design. We could just as well replace the cysteine and make a continuous ~ 70 amino acid protein that should recognize a predictable site" (column 2, page 9095). Applicant respectfully points out that such statements cannot be considered in isolation but must be interpreted in their proper context, namely in light of *all* of the teachings in the prior art at the time of the invention. First, Applicant notes that Park et al. did not in fact make such a protein, or demonstrate its ability to bind to DNA. A mere statement that something could possibly be done does not provide a reasonable expectation that it can be. Furthermore, Applicant notes that contemporaneously with Park et al., other researchers undertook just such a study (Deng et al., *Proc. Natl. Acad. Sci. USA* 89:8572, 1992, **Exhibit B**). These researchers produced recombinant proteins comprised of the basic region of c-Jun (a leucine zipper protein that homodimerizes and also forms a heterodimer with v-Jun), a short peptide loop, and a minimally-modified version of the basic region of c-Jun. Five of the six recombinant proteins that they made did not bind to DNA, and the only one that did bind did so with one-tenth the efficiency of a wild type c-Jun homodimer; five others did not bind (Figure 1, page 8573). Thus, one of ordinary skill in the art, considering Park et al. in the context of other available references, including Deng et al., would conclude that preparation of a *nucleic acid* that encodes a *protein* is undesirable, and chemical cross-linking of domains is preferred. Consistent with this, we note that in three subsequent papers published over the next four years, Park et al. continued to use the chemical cross-linking approach (Park et al., *Proc. Natl. Acad. Sci. USA* 90:4892, 1993; Park et al., *J. Am. Chem. Soc.* 117:6287, 1995; and Park et al., *J. Am. Chem. Soc.* 118:4235, 1996, copies of which are attached as **Exhibit C, D and E, respectively**).

None of the secondary references remedy the deficiencies of Park et al. In fact, both Mitchell et al. and Harrison focus strongly on the *differences* that exist between and among classes of DNA-binding domains. These references *teach away* from the idea that elements of different DNA-binding domains could be combined with one another and/or that zinc-finger domains will behave in the same way as basic regions from leucine-zipper proteins. Shultz et al. has no teaching or suggestion of chimeric nucleic acid-binding domains at all. Withdrawal of the rejection is therefore respectfully requested.

Rejections under 35 U.S.C. §103 over Gossen et al. in view of Park et al., Mitchell et al., Harrison and Schultz:

Claims 40-70 and 72-98 were rejected under 35 U.S.C. §103 as being unpatentable over Gossen et al. (U.S. Patent No. 5,464,758) in view of Park et al. (*Proc. Natl. Acad. Sci. USA* 89:9094, 1992), Mitchell et al. (*Science* 245:371, 1989), Harrison (*Nature* 353:715, 1991) and Schultz (*Nature* 240:426, 1988).

Claims 40-70 and 72-98 have been canceled. The cancelled claims have been replaced with new claims 99-119. Applicant respectfully traverses this rejection as applied to new claims 99-119.

Applicant respectfully submits that as set forth above, Park et al., Mitchell et al., Harrison and Schultz do not render obvious the claimed nucleic acids. Thus, even if, as contended by the Examiner, Gossen et al. does teach that "it is within the skill in the art to make a nucleic acid vector that encodes a chimeric transactivator fusion protein [...], make a nucleic acid encoding a heterologous protein operably linked to a regulator binding site that the chimeric protein binds to, place the nucleic acid in a eukaryotic cell, [...]" the cited references fail to provide motivation to combine the references and to provide the required reasonable expectation of success.

Furthermore, the entire goal of Gossen et al. is to provide tetracycline-responsive transcriptional regulators. Thus, any combination of Gossen et al. with another reference can only result in a teaching of a protein containing a tetR DNA binding domain, as it is this DNA binding domain that confers tetracycline-responsiveness. The tetR DNA binding domain is a helix-turn-helix domain. Applicant therefore submits that, even if there were some motivation to

make the combination of references suggested by the Examiner, the combination would only teach a protein containing two helix-turn-helix domains, at least one of which is the tetR domain; the combination would not teach or suggest the claimed invention.

Conclusion:

In light of the response filed April 2, 2003 (incorporated herein by reference), the foregoing Remarks and the Terminal Disclaimer filed herewith, Applicant respectfully submits that the present case is in condition for allowance. A Notice to that effect is respectfully requested. Applicant would like to take this opportunity to thank the Examiner for his careful consideration of this case. If it is believed that a telephone conversation would help expedite prosecution of this case, or if any further information is required, the Examiner is invited to contact the undersigned at (617) 248-4793. Additionally, please charge any fees that may be required, or credit any overpayment, to our Deposit Account No. 03-1721.

Respectfully submitted,

Charles Lyon, Ph.D.
Agent for Applicant
Limited Recognition Under 37 CFR §10.9(b)

CHOATE, HALL & STEWART
Exchange Place
53 State Street
Boston, MA 02109
(617) 248-5000

Dated: August 1, 2003

3589307_1.DOC

Certificate of Mailing

I certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as First Class Mail in an envelope addressed to Assistant Commissioner for Patents, Washington, DC 20231.

August 1, 2003 — Signature
Date

Typed or Printed Name of person signing certificate

22. W. B. Curry, R. C. Thunell, S. Honjo, *Earth Planet. Sci. Lett.* 64, 33 (1983); R. G. Fairbanks, M. Sverdlove, R. Free, P. H. Wiebe, A. W. H. Be, *Nature* 298, 841 (1982); W. G. Deuser, E. H. Ross, C. Hemleben, M. Spindler, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 33, 103 (1981).
23. A. Berger, *J. Atmos. Sci.* 35, 2362 (1978).

24. We thank W. F. Ruddiman and C. Sancetta for their critical reviews. This research was supported by NSF grants OCE89 11841 and OCE85-16133 Lamont-Doherty Geological Observatory Contribution No. 4628.

27 March 1990; accepted 4 June 1990

Sequence-Specific DNA Binding by a Short Peptide Dimer

ROBERT V. TALANIAN, C. JAMES MCKNIGHT, PETER S. KIM

A recently described class of DNA binding proteins is characterized by the "bZIP" motif, which consists of a basic region that contacts DNA and an adjacent "leucine zipper" that mediates protein dimerization. A peptide model for the basic region of the yeast transcriptional activator GCN4 has been developed in which the leucine zipper has been replaced by a disulfide bond. The 34-residue peptide dimer, but not the reduced monomer, binds DNA with nanomolar affinity at 4°C. DNA binding is sequence-specific as judged by deoxyribonuclease I footprinting. Circular dichroism spectroscopy suggests that the peptide adopts a helical structure when bound to DNA. These results demonstrate directly that the GCN4 basic region is sufficient for sequence-specific DNA binding and suggest that a major function of the GCN4 leucine zipper is simply to mediate protein dimerization. Our approach provides a strategy for the design of short sequence-specific DNA binding peptides.

THE TRANSCRIPTIONAL ACTIVATOR GCN4 (*1*), which is responsible for the general control of amino acid biosynthesis in yeast (*2*), binds DNA through a structural motif common to several proteins (*3*), including the nuclear oncogene products Fos and Jun. This "bZIP" (*4*) motif consists of a region with several basic residues that probably contacts DNA directly and an adjacent region of about 30 residues containing a heptad repeat of leucines, the "leucine zipper" (*5*), that mediates dimerization. Such bZIP dimers bind DNA sites that are approximately diad-symmetric (*3*).

Structural studies of a synthetic peptide corresponding to the leucine zipper region of GCN4 indicate that the peptide dimerizes as a parallel coiled coil (*6, 7*). The leucine zipper regions are necessary for dimerization of GCN4 (*8–10*) and other bZIP proteins (*11, 12*) and for heterodimer formation by the Fos and Jun proteins (*10, 13–15*). Moreover, synthetic leucine zipper peptides are sufficient for specific homodimer (*6*) and heterodimer (*16*) formation.

The basic region of bZIP proteins is important for DNA binding. Several bZIP proteins with mutations in the basic region fail to bind DNA sequence-specifically although they can dimerize (*12, 13, 17*). Alignment of sequences from different bZIP proteins shows that conserved residues in the basic region and the leucine zipper are separated by an invariant number of residues (*4, 18*). This separation appears crucial since insertion or deletion of a few amino acid residues at the boundary between the two regions can eliminate specific DNA binding activity (*14, 19, 20*). Nevertheless, the two regions appear capable of functioning autonomously, since chimeric bZIP domains (combining the basic region of one protein with the leucine zipper of another) often retain specific DNA binding activity (*10, 20, 21*).

We asked whether the basic region alone, dimerized with a disulfide in place of the leucine zipper, retains sequence-specific DNA binding activity. A peptide (GCN4-brl), corresponding to residues 222 to 252 of GCN4 (*22*), was synthesized (*23*) with a Gly-Gly-Cys linker (*6*) added at the carboxyl terminus (Fig. 1). The glycines were included to provide a flexible linker in the disulfide-bonded dimer, referred to as GCN4-brl^{ss}. The peptide was made as the carboxyl-terminal amide to avoid introduction of additional charge. A second peptide (GCN4-bZIP1), corresponding to the entire bZIP region of GCN4 (residues 222 to 281), was also synthesized (Fig. 1). This 60-residue peptide is capable of dimerization and sequence-specific DNA binding (*8*).

Fig. 2. Gel mobility shift assays (*24*) indicate that DNA binding by GCN4-brl^{ss}, but not GCN4-bZIP1, is sensitive to DTT. Lane 1, no peptide; lane 2, GCN4-brl^{ss}; lane 3, GCN4-brl^{ss} with 10 mM DTT; lane 4, GCN4-bZIP1; and lane 5, GCN4-bZIP1 with 10 mM DTT.

Gel mobility shift assays (*24*) indicate (Fig. 2) that both GCN4-brl^{ss} and GCN4-bZIP1 bind a 20-bp oligonucleotide, GRE20 (*24*), which contains the GCN4 recognition element (GRE) 5′-ATGACTCAT-3′ (*25*). As measured by titration of the gel shift, GCN4-brl^{ss} binds GRE20 with a dissociation constant of ~10 nM at 4°C. Reduction of the disulfide bond in GCN4-brl^{ss} by addition of 10 mM dithiothreitol (DTT) decreases substantially the amount of mobility-shifted DNA, whereas DNA binding by GCN4-bZIP1 is unaffected by this treatment (Fig. 2).

The DNA binding specificities of GCN4-brl^{ss} and GCN4-bZIP1 were tested by using deoxyribonuclease (DNase) I footprinting (*26*). At 4°C both peptides show sequence-specific protection of the GRE site from DNase I digestion (Fig. 3). However, when DNase I digestion was carried out at 24°C, GCN4-brl^{ss} failed to bind specifically, although GCN4-bZIP1 gave an identical footprint to that obtained at 4°C.

The DNA binding specificity of GCN4-brl^{ss} suggests that the peptide is a valid model for the DNA binding activity of GCN4. The binding activity of the peptide dimer demonstrates directly that the basic region of GCN4 (and presumably other bZIP proteins) contains sufficient information for sequence-specific DNA binding. The successful substitution of the leucine

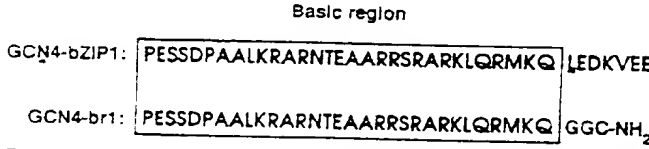

Fig. 1. Sequences of the peptides studied (*23*). GCN4-bZIP1 consists of the 60 carboxyl-terminal residues of GCN4 (*22*). The leucines in the leucine repeat are underlined. GCN4-brl consists of the basic region residues (boxed) plus the carboxyl-terminal linker Gly-Gly-Cys. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; G, Gly; H, His; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; and Y, Tyr.

Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142, and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

zipper with a flexible disulfide linker, and the dependence of DNA binding on the presence of the disulfide bond, suggest that the primary function of the leucine zipper is dimerization. However, DNA binding by GCN4-brl[55], but not GCN4-bZIP1, is temperature dependent between 4° and 24°C. These observations suggest an additional role for the leucine zipper [for example, orientation of the DNA binding regions; see (*14, 19, 20*)] that is not modeled by the flexible disulfide linker.

Structural studies can be simplified by using peptide models for protein motifs. Accordingly, we have used circular dichroism (CD) spectroscopy to examine the secondary structure of GCN4-brl[55] in the presence and absence of GRE20. The CD spectrum of the peptide (Fig. 4A) suggests that it shows partial α-helix formation in the absence of DNA (*27*). The intensity of the CD signal of GCN4-brl[55] at 222 nm (a helical band) increases substantially upon addition of an equimolar amount of GRE20 (Fig. 4B). The small change in the region of the spectrum dominated by signals from the oligonucleotide (245 to 310 nm) suggests that the much larger changes observed below 245 nm result primarily from changes in peptide rather than oligonucleotide structure. The difference spectrum (Fig. 4C) indicates that the peptide is highly α helical when bound to DNA (*28*). These results are consistent with both the "scissors grip" (*4*) and "induced helical fork" (*29*) models, which postulate that the basic regions of bZIP proteins bind DNA in an α-helical conformation.

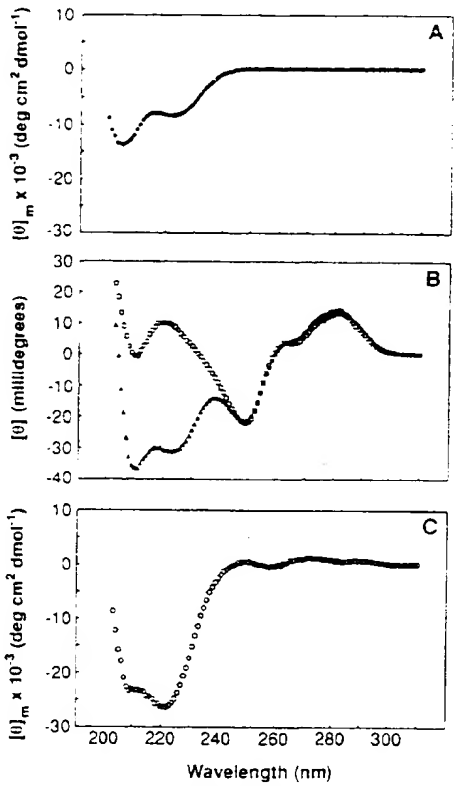

Fig. 4. CD difference spectroscopy indicates that GCN4-brl[55] is helical when bound to DNA (*32*). (**A**) GCN4-brl[55] alone. (**B**) GRE20 alone (□) and GCN4-brl[55] with GRE20 (▲). (**C**) Spectrum of GCN4-brl[55] bound to GRE20 calculated as the difference between the two spectra in (B).

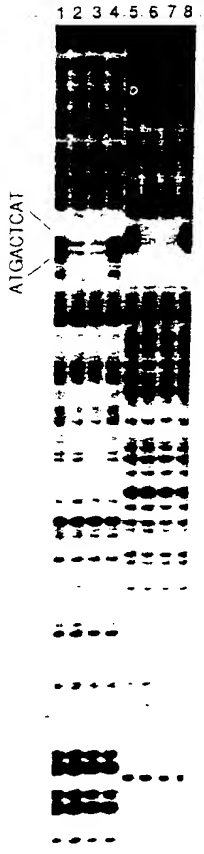

Fig. 3. The DNase I footprint of the GCN4 binding site by GCN4-brl[55] and GCN4-bZIP1 are identical (*26*). Lanes 1 to 4, DNA labeled with ^{32}P-phosphate at the 5′ end of the ⊖ strand; lanes 5 to 8, label on the ⊖ strand; lanes 1, 4, 5, and 8, DNase I control (no peptides present); lanes 2 and 6, GCN4-brl[55]; and lanes 3 and 7, GCN4-bZIP1.

Although GCN4-brl[55] is a remarkably short DNA binding peptide, it seems likely that even shorter peptides with sequence-specific DNA binding activity can be made. For example, several of the amino-terminal residues in the basic region used here have been found recently to be dispensable for DNA binding (*19, 30*). In addition, the use of a Gly-Gly-Cys (*6*) or other linker [see, for example, (*31*)] could lead to peptide models for other DNA binding motifs. Peptide models like GCN4-brl[55] hold promise for structural studies of sequence-specific protein-DNA interactions and for the design of short, sequence-specific DNA binding peptides.

REFERENCES AND NOTES

1. M. D. Penn, B. Galgoci, H. Greer, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2704 (1983); A. G. Hinnebusch and G. R. Fink, *ibid.*, p. 5374.
2. E. W. Jones and G. R. Fink, in *The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression*, J. N. Strathern, E. W. Jones, J. R. Broach, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982), pp. 181–299.
3. Reviewed by P. F. Johnson and S. L. McKnight, *Annu. Rev. Biochem.* **58**, 799 (1989); K. Struhl, *Trends Biochem. Sci.* **14**, 137 (1989).
4. C. R. Vinson, P. B. Sigler, S. L. McKnight, *Science* **246**, 911 (1989).
5. W. H. Landschultz, P. F. Johnson, S. L. McKnight, *ibid.* **240**, 1759 (1988).
6. E. K. O'Shea, R. Rutkowski, P. S. Kim, *ibid.* **243**, 538 (1989).
7. T. G. Oas, L. P. McIntosh, E. K. O'Shea, F. W. Dahlquist, P. S. Kim, *Biochemistry* **29**, 2891 (1990); R. Rasmussen, D. Benvegnu, E. K. O'Shea, P. S. Kim, T. Alber, unpublished results.
8. I. A. Hope and K. Struhl, *Cell* **46**, 885 (1986).
9. ———, *EMBO J.* **6**, 2781 (1987).
10. T. Kouzarides and E. Ziff, *Nature* **340**, 568 (1989); J. W. Sellers and K. Struhl, *ibid.* **341**, 74 (1989).
11. C. V. Dang, M. McGuire, M. Buckmire, W. M. F. Lee, *ibid.* **337**, 664 (1989).
12. W. H. Landschultz, P. F. Johnson, S. L. McKnight, *Science* **243**, 1681 (1989); V. J. Dwarki, M. Montminy, I. M. Verma, *EMBO J.* **9**, 225 (1990).
13. R. Turner and R. Tjian, *Science* **243**, 1689 (1989); R. Gentz, F. J. Rauscher III, C. Abate, T. Curran, *ibid.*, p. 1695; M. Neuberg, M. Schuermann, J. B. Hunter, R. Müller, *Nature* **338**, 589 (1989).
14. T. Smeal, P. Angel, J. Meek, M. Karin, *Genes Dev.* **3**, 2019 (1989).
15. L. J. Ransone, J. Visvander, P. Sassone-Corsi, I. M. Verma, *ibid.*, p. 770; M. Schuermann *et al.*, *Cell* **56**, 507 (1989).
16. E. K. O'Shea, R. Rutkowski, W. F. Stafford III, P. S. Kim, *Science* **245**, 646 (1989).
17. D. Bohmann and R. Tjian, *Cell* **59**, 709 (1989).
18. P. K. Vogt, T. J. Bos, R. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3316 (1987).
19. K. Struhl, personal communication.
20. P. Agre *et al.*, *Science* **246**, 922 (1989).
21. K. Struhl, *Cell* **50**, 841 (1987); Y. Nakabeppu and D. Nathans, *EMBO J.* **8**, 3833 (1989); M. Neuberg, J. Adamkiewicz, J. B. Hunter, R. Müller, *Nature* **341**, 243 (1989).
22. G. Thireos, M. D. Penn, H. Greer, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5096 (1984); A. G. Hinnebusch, *ibid.*, p. 6442.
23. Peptides were synthesized on an Applied Biosystems Model 430A peptide synthesizer with standard reaction cycles modified to include acetic anhydride capping. Peptides were cleaved from the resins by low-high HF cleavage (Immunodynamics, Inc., San Diego, CA) and desalted by Sephadex G-10 chromatography in 5% acetic acid. Purifications were by high-performance liquid chromatography with a Vydac reverse-phase C_{18} column and a linear gradient of CH_3 CN-H_2O with 0.1% trifluoroacetic acid. Fast atom bombardment mass spectrometry: GCN4-brl: calculated, 3796.5; found, 3795.8; GCN4-bZIP1: calculated, 7015.4; found, 7015.5.
24. M. M. Garner and A. Revzin, *Nucleic Acids Res.* **9**, 3047 (1981); M. Fried and D. M. Crothers, *ibid.*, p. 6505. Binding solutions contained in 15 μl: 20 mM tris (pH 7.4), 4 mM KCl, 2 mM $MgCl_2$, 2 mM EDTA, 0.1% NP-40 detergent, 5000 cpm (~2 fmol) 5′-^{32}P-labeled GRE20 (5′-CGGATGACTCATTTTTTTTC-3′, double-stranded) and 250 nM peptides when present. Mixtures were incubated at 4°C for 30 min. Free and peptide-bound labeled GRE20 were resolved by nondenaturing 8% polyacrylamide gel electrophoresis in TE buffer at 4°C. Peptide concentrations were determined by tyrosine absorbance [H. Edelhoch, *Biochemistry* **6**, 1948 (1967)] or by quantitative amino acid analysis, and DNA concentration was determined with an extinction coefficient at 260 nm (ϵ_{260}) of 473,000 M^{-1} cm^{-1} [R. B. Wallace and C. G. Miyada, *Methods Enzymol.* **152**, 423 (1987)].
25. D. E. Hill, I. A. Hope, J. P. Macke, K. Struhl, *Science* **234**, 451 (1986).
26. D. J. Galas and A. Schmitz, *Nucleic Acids Res.* **5**, 3157 (1978). Solutions (4°C) contained in 200 μl: 20 mM tris (pH 7.4), 4 mM KCl, 2 mM $MgCl_2$, 0.1% NP-40, 200 μM (bp) sonicated salmon sperm DNA, 15,000 cpm (~5 fmol) of DNA probe and 1.0 μM GCN4-brl[55] and 1.1 μM GCN4-bZIP1 when present. DNA probes were prepared by poly-

merase chain reaction (PCR) amplification of pUC9-Sc4251 (*33*) with synthetic 5′-^{32}P–labeled or unlabeled 17-residue primers defining a 231-bp PCR product with the GRE centered. Nuclease digestion (90 s at 4°C) was initiated by addition of 0.2 μg DNase I (Sigma) and $CaCl_2$ to 2.5 mM and was quenched by addition of 200 μl of 1% SDS, 200 mM NaCl, 20 mM EDTA, and yeast transfer RNA (25 μg/ml) (Sigma). Samples were purified by phenol-chloroform extraction and ethanol precipitation and were run on a 6% sequencing (7.7 M urea) polyacrylamide gel.

27. α-Helices can be inferred from CD spectra obtained in aqueous solution with much higher confidence than for other secondary structures, although strictly speaking we cannot distinguish between α- and 3_{10}-helices; see R. W. Woody, in *The Peptides*, S. Undenfriend, J. Meienhofer, J. R. Hruby, Eds. (Academic Press, New York, 1985), vol. 7, pp. 15–114; Y.-H. Chen, J. T. Yang, K. H. Chau, *Biochemistry* 13, 3350 (1974).
28. CD experiments suggest that the helical content in GCN4-bZIP1 also increases substantially upon binding GRE20 (A. D. Frankel, E. K. O'Shea, T. G. Oas, P. S. Kim, unpublished results). These experiments are difficult to interpret, however, because preliminary two-dimensional nuclear magnetic resonance studies (L. P. McIntosh, T. G. Oas, P. S. Kim, unpublished results) indicate that the leucine zipper region of GCN4-bZIP1 is substantially less stable than the isolated leucine zipper (*6*, *7*).
29. K. T. O'Neil, R. H. Hoess, W. F. DeGrado, *Science* 249, 774 (1990).
30. M. G. Oakley and P. B. Dervan, *Science* 248, 847 (1990).
31. T. G. Oas and P. S. Kim, *Nature* 336, 42 (1988); J. P. Staley and P. S. Kim, *ibid.* 344, 685 (1990).
32. CD spectra were obtained with an AVIV model 60HDS CD spectrometer at 25°C in a 5-mm cell. Samples contained 10 mM phosphate buffer (pH 7.0), 100 mM NaCl, and 4.6 μM GCN4-bzi1 and 5.0 μM GRE20 when present. Spectra in (A) and (B) were the average of multiple scans and were baseline-corrected with a spectrum of buffer alone, but were not smoothed.
33. Plasmid pUC9-Sc4251, containing the GRE sequence (*25*), has the 1.3-kb Eco RI–Bam HI fragment of plasmid YIp55-Sc4251 (*25*) cloned into the Eco RI–Bam HI site of pUC9 and was kindly provided by K. Struhl.
34. We thank A. Frankel for advice and discussions in all aspects of this work, E. O'Shea for preliminary experiments and discussions, R. Rutkowski for expert peptide synthesis, and S. Stradley and L. Gierasch for performing quantitative amino acid analysis. Supported by National Research Service Award GM13665 from the National Institutes of Health (R.V.T.), a postdoctoral fellowship from the Massachusetts Division of the American Cancer Society (C.J.M.), and by grants from the National Institutes of Health (GM44162) and the Lucille P. Markey Charitable Trust (P.S.K.).

1 May 1990; accepted 22 June 1990

Evidence of Changes in Protease Sensitivity and Subunit Exchange Rate on DNA Binding by C/EBP

JON D. SHUMAN, CHARLES R. VINSON, STEVEN L. MCKNIGHT

The transcription factor C/EBP uses a bipartite structural motif to bind DNA. Two protein chains dimerize through a set of amphipathic α helices termed the leucine zipper. Highly basic polypeptide regions emerge from the zipper to form a linked set of DNA contact surfaces. In the recently proposed a "scissors grip" model, the paired set of basic regions begin DNA contact at a central point and track in opposite directions along the major groove, forming a molecular clamp around DNA. This model predicts that C/EBP must undertake significant changes in protein conformation as it binds and releases DNA. The basic region of ligand-free C/EBP is highly sensitive to protease digestion. Pronounced resistance to proteolysis occurred when C/EBP associated with its specific DNA substrate. Sequencing of discrete proteolytic fragments showed that prominent sites for proteolysis occur at two junction points predicted by the "scissors grip" model. One junction corresponds to the cleft where the basic regions emerge from the leucine zipper. The other corresponds to a localized nonhelical segment that has been hypothesized to contain an N-cap and facilitate the sharp angulation necessary for the basic region to track continuously in the major groove of DNA.

THE TRANSCRIPTION FACTOR C/EBP regulates gene expression in a variety of tissues, including liver, adipose, lung, and intestine. The protein binds DNA through a bipartite structural motif consisting of a dimer-forming region immediately preceded by a polypeptide region rich in basic amino acids. Leucine residues occur in a heptad array along the dimer interface. Anticipating that the leucine residues would provide attractive, intersubunit interactions, we termed the dimer-forming region the leucine zipper (*1*). Biophysical studies have documented the α-helical nature of the leucine zipper and have shown that helices intertwine around one another in a parallel orientation (*2*). Considerable evidence has confirmed the role of the leucine zipper in dimerization of both identical and nonidentical protein subunits (*3*).

A variety of observations on transcription factors of this class have indicated that direct contact with DNA is mediated by the basic region. For example, a chimeric protein containing the basic region of C/EBP linked to the leucine zipper of GCN4 binds DNA with the specificity of C/EBP (*4*).

Proteins that use the contiguous basic region–leucine zipper arrangement (bZIP proteins) exhibit an invariant, six–amino acid spacing between the two components. Noting this fixed spatial register, as well as an absence of Pro and Gly residues, Vinson and colleagues (*5*) predicted that the basic region, like the zipper, would adopt an α-helical conformation. DNA-bound protein was hypothesized to form a Y-shaped molecule, the stem and arms corresponding, respectively, to paired zippers and bifurcating basic regions. This arrangement allowed the two basic regions to penetrate the major groove of DNA from a common point (the cleft of the Y), then track in opposite directions along each half of a dyad-symmetric binding site. Finally, this modeling predicted that α-helical structure would be locally disrupted within the basic region, facilitating a sharp bend necessary to allow continuous tracking of each basic region around the DNA on the side opposite to initial entry.

This model for bZIP proteins has been compared to the "scissors grip" hold that a wrestler uses to grasp the torso of an opponent. By wrapping around the DNA molecule on the side opposite of initial entry, the two subunits of a bZIP protein form a molecular clamp. If correct, this model demands that the protein undertake significant conformational changes as it binds and releases DNA. It further predicts that subunit exchange, which occurs rapidly in the absence of DNA, should be slowed dramatically upon DNA binding.

We examined the susceptibility of C/EBP to trypsin cleavage in the presence and absence of its DNA substrate. Trypsin, which cleaves the peptide bond carboxyl terminal to Arg and Lys residues, is a sensitive probe of the folded state (*6*). Moreover, C/EBP contains eight potential sites for trypsin cleavage in its basic region, six in its leucine zipper, and two in the short segment that links the basic region to the zipper (Fig. 1A).

Purified C/EBP (*7*) was exposed for 1-min intervals to varying amounts of trypsin. Digestion products were separated by electrophoresis on an SDS-polyacrylamide gel, transferred to nitrocellulose, and detected by immunoblotting with an antibody (α-C) specific to the carboxyl terminus of C/EBP (*8*). This strategy (*9*) provided a fixed labeling site on C/EBP, thus allowing a reasonably accurate identification of the sites of trypsin cleavage.

The patterns of trypsin cleavage of C/EBP alone, or of protein samples that had been mixed with either nonspecific or specific

Howard Hughes Research Laboratories, Department of Embryology, Carnegie Institution of Washington, Baltimore, MD 21210.

Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 8572–8576, September 1992
Genetics

Construction and expression of a monomeric c-Jun protein that binds and activates transcription of AP-1-responsive genes

(DNA-binding protein/transactivation)

TILIANG DENG AND MICHAEL KARIN

Department of Pharmacology, Center for Molecular Genetics, University of California, San Diego, School of Medicine, 0636, 9500 Gilman Drive, La Jolla, CA 92093

Communicated by Mark Ptashne, April 24, 1992

ABSTRACT **c-Jun is a typical member of the bZIP (basic zipper) family of dimeric transcriptional activators. These proteins contain a basic region responsible for DNA sequence recognition and a leucine zipper that mediates dimerization. bZIP proteins regulate a large number of important physiological functions and, therefore, present an interesting target for molecular interference and mimicry. As a step toward the development of peptide and nonpeptide analogs of such proteins, we constructed a derivative of c-Jun that binds DNA as a monomer. This construction was done by connecting a second basic region to the natural basic region of c-Jun by means of a short peptide loop. Although the polypeptide backbone of the second basic region has an inverted polarity relative to that of the natural basic region, the monomeric c-Jun protein binds DNA with reasonably high affinity and indistinguishable specificity from the wild-type, dimeric c-Jun protein. Furthermore, the monomeric c-Jun protein can activate transcription *in vivo*. These findings indicate that the polypeptide backbone of the basic region contributes little to sequence recognition and that the leucine zipper is not directly involved in transcriptional activation.**

Many sequence-specific transcription factors, both prokaryotic and eukaryotic, interact with DNA as preformed dimers (1–8). Two large families of dimeric eukaryotic transcription factors were recently identified: the bZIP (for basic zipper) and the HLH (helix–loop–helix) proteins (3–5, 7, 8). These proteins are involved in a variety of physiological functions, including the control of cell proliferation and differentiation and in mediating the actions of polypeptide hormones, cytokines, and growth factors. The DNA-binding domains of both families are constructed of a basic region rich in positively charged amino acids, which interacts directly with the DNA, and an adjacent dimerization motif. The bZIP dimerization motif is an amphipathic α-helix containing several heptad repeats of leucine residues, responsible for formation of a parallel coiled-coil known as the leucine zipper (3, 9, 10). In both cases, the dimerization domains mediate not only homotypic interactions but also heterotypic interactions that expand the regulatory potential of these proteins. For example, a c-Jun–c-Fos heterodimer is more stable than a c-Jun homodimer and, therefore, has higher DNA-binding activity and is a more efficient transcriptional activator (11–17). Heterodimerization of MyoD or myogenin with E12 and E47 increases their affinity to the E box sequence of muscle-specific promoters (5, 7, 8).

The localization of dimerization and DNA-binding functions of bZIP and helix–loop–helix proteins to relatively small and well-defined sequence motifs has raised the possibility of synthesizing analogs of these proteins that could interfere with either their dimerization or DNA-binding activities. Indeed, several groups have described that short synthetic peptides corresponding in sequence to the basic regions and leucine zippers of certain bZIP proteins can bind DNA *in vitro* (18, 19). We are interested in preparing analogs of c-Jun that are functional *in vivo* and could be prototypes for designing totally synthetic analogs; these synthetic analogs could eventually be used as competitive inhibitors of DNA binding. We also wanted to determine whether the leucine zipper of c-Jun is required for any other activity besides dimerization. By constructing a c-Jun protein that binds DNA as a monomer, we show that dimerization is not essential for transcriptional activation and that c-Jun can activate transcription by itself, without forming dimers with other bZIP proteins.

MATERIALS AND METHODS

Plasmids, Cell Culture, and Transfections. Construction of c-Jun, cJunΔLZ expression vectors, and the $-79/+170$jun–CAT, $-79/+170\Delta$AP-1jun–CAT reporters has been described (20–22). To generate the monomeric c-Jun expression vectors, codons 278 and 279 of c-Jun in the Rous sarcoma virus–c-Jun vector (20) were mutated from GCC CGG to GCG CGC to create a *Bss*HII site. The resulting plasmid was digested by *Bss*HII and *Xho* I and ligated to phosphorylated oligonucleotides coding for the loop and a new basic region as shown in Fig. 1*C*. The exact sequences of the oligonucleotides are available upon request. To construct the truncated Jun (t-Jun) expression vector a *Pst* I–*Bam*HI fragment encoding amino acids 222–331 of c-Jun was cloned into pET-8C (23) by using the adaptor: 5′–CATGGCTAGCGAATTCCTGCA
3′ –CGATCGCTTAAGG–5′.

F9 cells were grown and transfected as described (20, 21).

Expression and Purification of Recombinant Proteins. To adapt the c-Jun cDNA to the pET-8c vector (23), two nucleotides preceding its initiator ATG codon were mutated to create a *Bsp*HI site by site-directed mutagenesis. The *Bsp*HI–*Bam*HI fragment from Rous sarcoma virus–c-Jun (20) was inserted into pET-8c between the *Nco* I and *Bam*HI sites to generate pET-8c/c-Jun. To express monomeric Jun (m-Jun), the C-terminal coding region of c-Jun in pET-8c/c-Jun was replaced by the same region of m-Jun. The plasmids were transformed into *Escherichia coli* BL21(DE3)pLysS. The cells were induced, and Jun proteins were extracted from inclusion bodies and renatured as described (24). The proteins used in this report were purified to near homogeneity by heparin–agarose chromatography (24). Protein concentrations were determined by the Bradford assay (Bio-Rad). The N-terminal sequence of the recombinant c-Jun was determined by J. Woodgett (Ludwig Institute for Cancer Research) as NH_2-Thr-Ala-Lys-Met-Glu-Thr-Thr, the expected sequence after removal of the first methionine residue. Trans-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: TRE, phorbol 12-myristate 13-acetate-responsive element; DSS, disuccinimidyl suberate; t-Jun, truncated Jun; m-Jun, monomeric Jun.

EXHIBIT B

fection and immunoprecipitation of protein in F9 cells were done as described (25, 26).

Mobility-Shift Assay. Mobility-shift assays (27) contained the indicated amounts of the different Jun proteins, 1 ng of ^{32}P-labeled phorbol 12-myristate 13-acetate response element (TRE) probe, 100 ng of sonicated salmon sperm DNA, 12 mM Hepes·KOH (pH 8.0), 50 mM KCl, 6 mM $MgCl_2$, 1 mM EDTA, 10% (vol/vol) glycerol, 5 mM dithiothreitol, and 80 μg of bovine serum albumin in a total volume of 20 μl. After a 20-min incubation at room temperature, reaction mixtures were loaded on 5% native polyacrylamide gels (acrylamide/bisacrylamide, 40:1). Electrophoresis was done in 0.4× Tris/borate/saline (TBE; 1× TBE is 90 mM Tris/64.6 M boric acid/2.5 mM EDTA, pH 8.3) at room temperature. The mobility-shift experiments were quantitated by counting the radioactivity of the dried gels with the Ambis radioanalytic imaging system.

jun–TRE, consensus TRE (16), NF1, and Sp1 (28) oligonucleotide probes were described previously.

DNase I Footprinting and Methylation Interference. The c-*jun* promoter probe was labeled at the *Nco* I site at position −132 on the noncoding strand (20) and incubated with either c-Jun (80 ng), m-Jun (1.6 μg), or bovine serum albumin (10 μg) and digested with either 1 or 3 ng of DNase I, as described (29). For methylation interference the c-*jun* promoter fragment (−132 to +170) was labeled at position −132 either on the coding (by T4 polynucleotide kinase) or noncoding (by avian myeloblastosis virus reverse transcriptase) strands. Methylation interference was done as described (29).

Chemical and UV Cross-Linking and Sedimentation Analysis. One hundred microliters of either c-Jun or m-Jun (both at 0.07 mg/ml) were treated with either 2 μl of dimethyl sulfoxide or 2 μl of 10 mM disuccinimidyl suberate (DSS) in dimethyl sulfoxide for 10 min at room temperature. The reactions were quenched by adding 5 μl of 1 M lysine and analyzed by electrophoresis on a 10% polyacrylamide/SDS gel stained with Coomassie blue. For UV cross-linking experiments, protein–DNA complexes were allowed to form for 20 min on ice. Samples were treated with UV light (254 nm) for another 20 min on ice, 4 cm from the light source. For further cross-linking by DSS, 2 μl of 10 mM DSS was added to each sample. The mixture was incubated at room temperature for 10 min and then quenched by adding 2 μl of 1 M lysine. After this, samples were boiled in Laemmli sample buffer and analyzed on SDS/12% PAGE. The gel was dried and exposed with intensifying screen at −80°C overnight.

Two micrograms of purified c-Jun or m-Jun was mixed with protein molecular mass markers (Bio-Rad) and sedimented through a 15–60% (vol/vol) glycerol gradient in buffer Z [25 mM Hepes·KOH, pH 8.0/12.5 mM $MgCl_2$/10% (vol/vol) glycerol/0.1% Nonidet P-40 1 mM dithiothreitol] containing 100 mM KCl. After 19 hr at 50,000 rpm in an SW55.2 rotor at 20°C the gradient was fractionated, and each fraction was analyzed by SDS/PAGE, silver staining, and immunoblotting for the presence of the molecular mass markers, c-Jun and m-Jun.

RESULTS

Experimental Approach. c-Jun is a bZIP protein that is a major component of the AP-1 complex, consisting of Jun homo- and heterodimers and Jun–Fos heterodimers (30, 31). These proteins interact with a common sequence known as the AP-1 site or the TRE (30, 31). Like other bZIP proteins, the leucine zipper of c-Jun determines its ability to form homo- and heterodimers (14–17). The basic region appears unstructured before DNA binding and assumes a helical conformation after contacting its recognition site (18, 19, 32, 33). From these and other findings, Vinson *et al.* (34) proposed that upon interaction with the DNA, the basic region undergoes structural transition, allowing the protein to bind

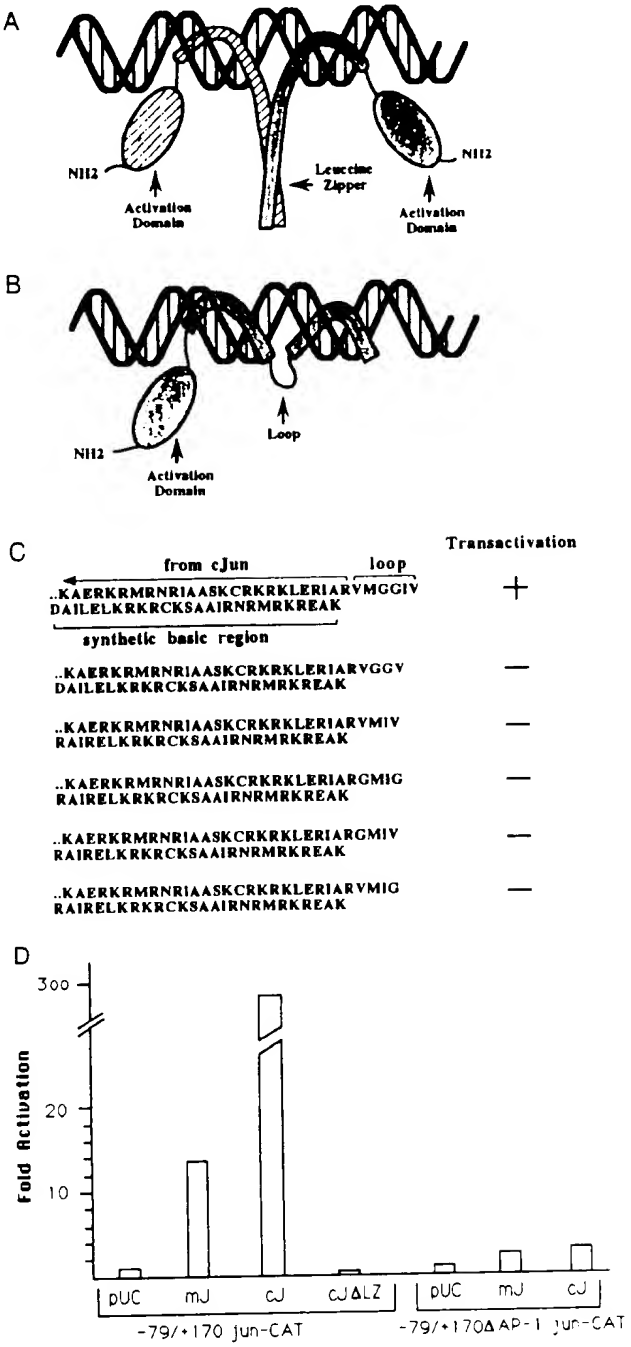

FIG. 1. Schematic representation of wild-type c-Jun (*A*), according to the scissors-grip model and monomeric c-Jun (*B*). (*C*) Primary structures of the DNA-binding domains of the monomeric c-Jun proteins using the single-letter amino acid code; the loop sequences are indicated in italics. Note that the second basic region has been modified to contain leucine and aspartate, instead of the last two arginines in the original c-Jun sequence. These arginines are not conserved among other bZIP proteins (18). In addition, the polypeptide backbone of the second basic region is in inverted polarity to that of the original basic region. Ability of the different monomeric c-Jun constructs to activate the c-*jun* promoter is indicated as positive (+) or negative (−). (*D*) Transactivation by c-Jun (cJ), m-Jun (mJ), and cJunΔLZ. F9 cells were transfected with the indicated reporters and expression vectors (2 μg of each plasmid per plate), and chloramphenicol acetyltransferase (CAT) activity was determined 24 hr later. The results are the mean of three experiments and are presented as the fold increase in acetyltransferase activity over the base line seen with cJunΔLZ.

its cognate DNA sites like a scissors grip (Fig. 1*A*). According to this model, it may be possible to link two basic regions by a peptide loop, instead of a leucine zipper, to generate a bZIP protein that binds DNA as a monomer (Fig. 1*B*). Hence, we connected a second, slightly modified, basic region to the basic region of c-Jun via the peptide-loop sequences shown in Fig. 1*C*. Glycines were included to increase loop flexibility. To allow synthesis of the protein as a single polypeptide chain, the second basic region has the same amino acid sequence as the first region, but this sequence follows the C-terminal to N-terminal direction. Despite its inverted polarity, the second basic region displays the same order of side chains as the original basic region, and if the peptide backbone itself does not participate in DNA binding, it may possess similar DNA-binding specificity. To identify a construct encoding a potentially monomeric c-Jun protein capable of functioning *in vivo*, we left the transactivation domain as part of the protein because, even though this domain is not necessary for DNA binding, it helps monitor activity of the protein. The various constructs were tested for their ability to transactivate the AP-1-responsive c-*jun* promoter (20). One construct tested was functional (Fig. 1*D*). Because this construct displayed much lower activity toward a mutated c-*jun* promoter, lacking a functional AP-1 site, this construct apparently acted in a sequence-specific manner. Immunoprecipitation analysis of transfected F9 cells indicated that the monomeric c-Jun construct expressed a protein with the predicted mobility (Fig. 2). Expression of this protein was 8-fold less efficient than expression of wild-type c-Jun, probably due to the more rapid degradation of the monomeric protein. This decreased expression could account for much of the decreased transactivation potential of the monomeric c-Jun construct.

The Designed Protein Is Monomeric Before and After DNA Binding. To further characterize its activity and physical properties, the protein encoded by this construct, m-Jun, and its wild-type counterpart, c-Jun, were expressed in *E. coli* by using the T7 expression system (23). Both proteins were extensively purified, and their aggregation state was examined by chemical cross-linking and sedimentation analysis. Treatment of c-Jun with the homobifunctional cross-linking agent (DSS) resulted in the formation of stable c-Jun dimers, whereas no cross-linking of m-Jun was seen (Fig. 3*A*). Sedimentation analysis with glycerol gradients indicated that c-Jun exists in solution as a mixture of monomers and dimers, whereas m-Jun is exclusively monomeric (Fig. 3*B*).

To demonstrate unequivocally that m-Jun binds to the TRE as a monomer, we did additional cross-linking experiments. Both c-Jun and m-Jun were incubated with a large excess of ^{32}P-labeled *jun*–TRE sufficient to saturate both proteins. These mixtures were exposed to UV irradiation to cross-link the protein molecules to DNA and DSS to cross-link protein molecules to each other. In preliminary experiments we found that neither cross-linking agent alone was sufficient for generating a composite protein–protein and protein–DNA adduct. After cross-linking, the mixtures were resolved on polyacrylamide/SDS gels, and the protein–DNA adducts were visualized by autoradiography. Fig. 3*C* shows that the c-Jun–TRE adduct migrated with an apparent molecular mass of 96 kDa, consistent with binding of a protein dimer to the TRE. However, the m-Jun–TRE adduct migrated with an apparent molecular mass of 46 kDa, consistent with binding of a protein monomer to the TRE.

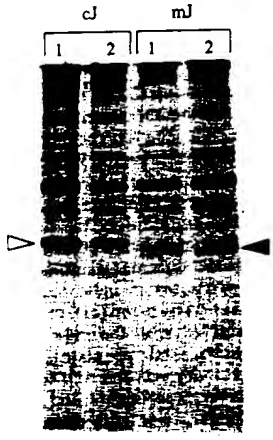

FIG. 2. Immunoprecipitation analysis of Jun protein expression. Expression vectors encoding wild-type c-Jun (cJ) and m-Jun (mJ) were transfected into F9 cells. Twelve hours after transfection the cultures were labeled for 3 hr, and ^{35}S-labeled Jun proteins were isolated by immunoprecipitation and resolved by PAGE. Migration positions of c-Jun and m-Jun are indicated by open and solid arrowheads, respectively. Two separate experiments are shown.

Monomeric c-Jun Binds DNA Specifically and Efficiently. Mobility-shift assays were done to compare the relative affinities of c-Jun and m-Jun to the *jun*–TRE (Fig. 4*A*). m-Jun was ≈one-tenth as efficient as c-Jun in binding this sequence. The complex formed by m-Jun with either the *jun*–TRE or a consensus TRE sequence had an electrophoretic mobility intermediate to those of the slower moving complex formed by wild-type c-Jun and the faster moving complex formed by a truncated c-Jun (t-Jun), consisting of its 110 C-terminal amino acids (Fig. 4*B*). These differences in electrophoretic mobility are consistent with m-Jun binding to the TRE as a monomeric 36-kDa protein, whereas c-Jun and t-Jun bind as dimeric 38-kDa and 15-kDa proteins, respectively. All three proteins bound both TRE probes with similar efficiencies, and competition experiments showed that binding of m-Jun to the *jun*–TRE was specific (Fig. 4*C*).

The specificity of m-Jun binding to DNA was further demonstrated by DNase I footprinting (Fig. 5*A*) and methylation interference (Fig. 5*B*). Both c-Jun and m-Jun generated indistinguishable protection and interference patterns centered around the TRE of the c-*jun* promoter. Interestingly, methylation of the first guanine upstream to the 5′-TGACATCA-3′ sequence fully interfered with binding of both c-Jun and m-Jun, whereas methylation of the second guanine partially interfered with their binding. Hence, both Jun proteins appear to contact these residues, even though they are not a part of the TRE core. These results, which are consistent with previous results obtained by mobility-shift assays (35), demonstrate that sequences that flank the TRE are also important for recognition by Jun proteins.

DISCUSSION

Collectively, these results indicate that m-Jun specifically recognizes the TRE *in vitro* and *in vivo*. Although it binds DNA as a monomer, m-Jun interacts with its recognition sites indistinguishably from c-Jun. These results are striking, considering the fact that the second basic region of m-Jun is polymerized in the C-terminal to N-terminal direction. These findings underscore the inherent flexibility of the basic region as a DNA-binding motif. A variety of experiments suggest that before DNA binding the basic region is unstructured but assumes a helical structure after DNA binding (18, 19, 32, 33). In addition to the structural transition of the basic region upon interaction with its target, the target sequence itself undergoes bending, resulting in even a better fit between the DNA and protein (36). Our results indicate that the polypeptide backbone of the basic region is not involved in sequence recognition. The polypeptide backbone does not appear directly involved in contacting the DNA in most other DNA-binding proteins, the structure of which has been determined at high resolution (37). However, we demonstrate that a DNA sequence-recognition motif can be polymerized in a polarity opposite to that of the natural structure and still maintain its activity and specificity. Even though m-Jun still contains one normal basic region that probably makes an important contribution to binding, the footprinting and methylation interference experiments indicate that both halves of

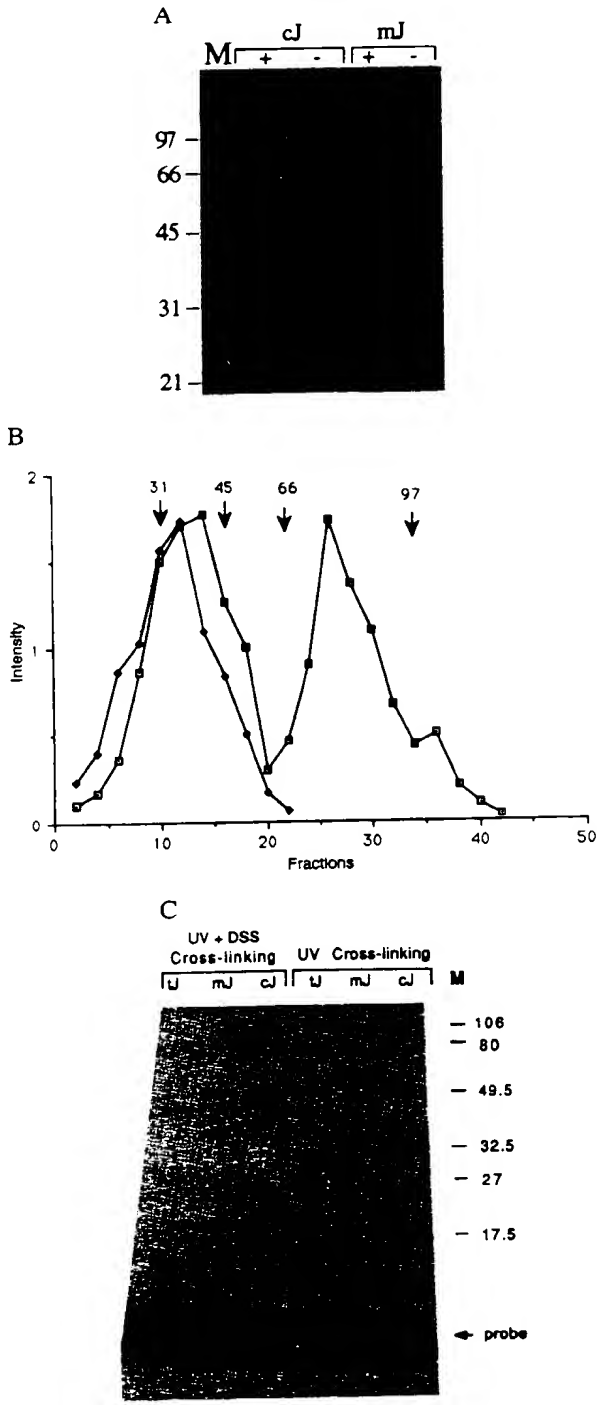

FIG. 3. m-Jun is a monomer in solution and upon binding to DNA. The aggregation state of m-Jun (mJ) was compared with that of c-Jun (cJ) by chemical cross-linking (*A*) and sedimentation (*B*) analyses. In *A* −, dimethyl sulfoxide; +, DSS/dimethyl sulfoxide. In *B* the graph shows relative concentrations of c-Jun (□) and m-Jun (◆) in different fractions, as determined by densitometry with an LKB UltroScan XL and peak positions of the molecular mass markers. (*C*) Aggregation state of protein–DNA complexes formed by m-Jun (mJ), t-Jun (tJ), and c-Jun (cJ) was analyzed by UV and DSS cross-linking. m-Jun (200 ng), t-Jun (40 ng), and c-Jun (40 ng) were incubated with ^{32}P-labeled *jun*–TRE probe (1 ng) for 20 min and then subjected to cross-linking by either UV alone or UV plus DSS. M, molecular mass markers.

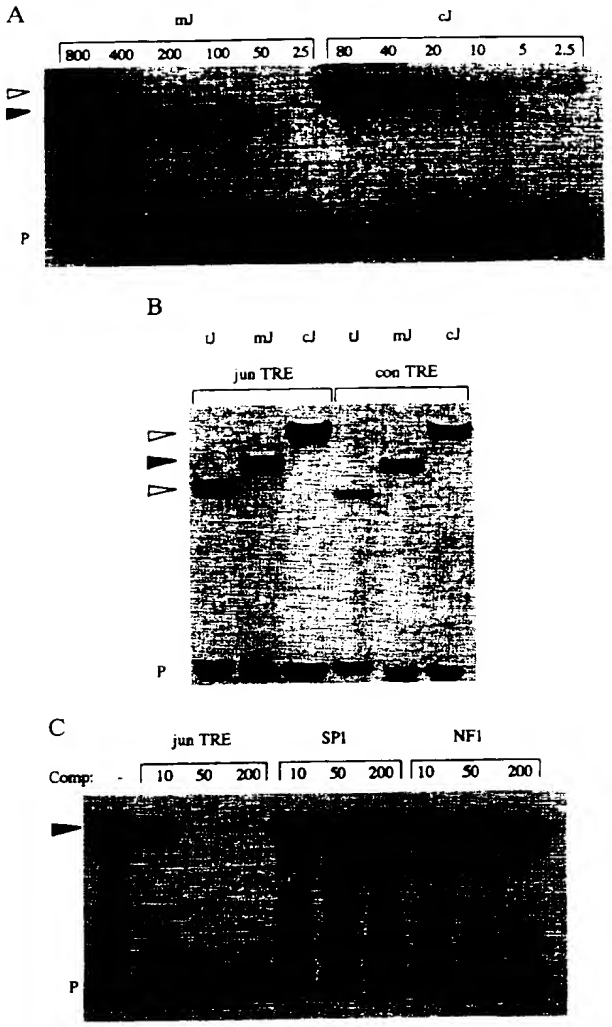

FIG. 4. Mobility-shift assays. (*A*) Protein titration experiment. A fixed amount (1 ng) of end-labeled *jun*–TRE probe (P) was incubated with increased amounts of c-Jun (cJ) and m-Jun (mJ), as indicated (in ng). Formation of protein–DNA complexes (solid arrow for m-Jun and open arrow for c-Jun) was analyzed by the mobility-shift assay. (*B*) Full-length c-Jun (cJ), truncated c-Jun (tJ), and m-Jun (mJ) were incubated with *jun*–TRE and consensus (con) TRE probes; the protein–DNA complexes (solid arrow for m-Jun, open arrows for c-Jun and t-Jun) were separated from free probes (P) by electrophoresis on a nondenaturing polyacrylamide gel. (*C*) Competition (Comp) experiment. m-Jun (400 ng) was incubated with 1 ng of the *jun*–TRE probe in the presence of the indicated amount (in ng) of unlabeled *jun*–TRE, Sp1, and NF1-binding-site oligonucleotides.

the TRE are contacted by the protein in a similar manner and to the same extent. Thus, it appears possible that as long as the basic region can project the same order of side chains into the major groove, it can bind DNA in a sequence-specific manner. These findings are encouraging for the future design of synthetic DNA-binding domains and suggest that such domains could be generated by anchoring appropriate side chains into a flexible polymeric backbone other than a polypeptide. The use of a nonpolypeptide backbone is likely to increase the biological half-life of the polymer, as it will not be recognized by cellular proteases.

Although transactivation by m-Jun was considerably lower than transactivation by c-Jun, immunoprecipitation indicates that m-Jun was also expressed less efficiently than c-Jun.

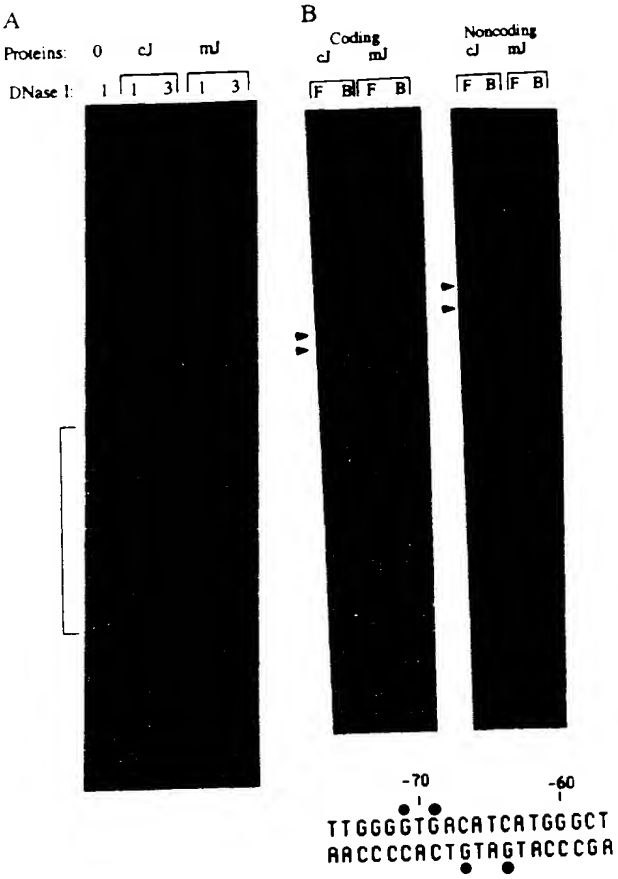

FIG. 5. DNase I footprinting (*A*) and methylation interference (*B*) analysis of c-Jun and m-Jun. Arrows indicate the guanine residues at which methylation strongly interfered with binding of c-Jun and m-Jun. Location of these guanines within the c-*jun* AP-1-binding site is indicated by circles at the bottom. The guanine residue that partially interfered with protein binding is not marked. F, free; B, bound.

Taking into consideration the 8-fold difference in the level of expression of the two proteins, m-Jun could function *in vivo* almost as efficiently as c-Jun. We noticed that another monomeric c-Jun construct with a loop only two amino acids shorter than m-Jun cannot transactivate the *jun* promoter (Fig. 1*C*). A small protein analogous to m-Jun was described by Talanian *et al.* (19), who connected two GCN4 (responsible for general control of amino acid biosynthesis in yeast) basic-region peptides via a disulfide bridge. Although this protein bound DNA at 4°C *in vitro*, it is unlikely that the disulfide bridge will remain oxidized at higher physiological temperatures and the reducing intracellular environment.

Our results strongly suggest that the only function of the leucine zipper is to mediate protein dimerization. As long as two basic regions can be tethered together at the right geometry, the leucine zipper is not required for either transactivation or for conferring binding specificity.

The approach described here can be used to assess the ability of other bZIP and probably also helix–loop–helix proteins to activate transcription by binding to their natural recognition sites. This is an important test because the ability of a given protein to activate transcription may depend on the binding-site type with which it interacts (5, 7, 8). Because these proteins will not be able to interact with other family members, this approach would reveal their intrinsic activity. Finally, the availability of monomeric derivatives of sequence-specific activators should simplify their structural analysis with nuclear magnetic resonance by alleviating problems associated with protein dimerization.

We thank Dr. J. Woodgett for N-terminal analysis of c-Jun, Drs. C. Murre and P. Angel for comments on the manuscript, Drs. W. DeGrado and M. Weiss for useful discussions, and Ms. K. Sexton for secretarial assistance. This work was supported by grants from the National Institutes of Health (ES04151, CA50528) and the Cancer Research Coordination Committee of the University of California. Tiliang Deng was supported, in part, by a fellowship from the Leukemia Society of America.

1. Takeda, Y., Ohlendorf, D. H., Anderson, W. F. & Matthews, B. W. (1983) *Science* **221**, 1020–1026.
2. Pabo, C. O. & Sauer, R. T. (1984) *Annu. Rev. Biochem.* **53**, 293–321.
3. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988) *Science* **240**, 1759–1764.
4. Abel, T. & Maniatis, T. (1989) *Nature (London)* **341**, 24–25.
5. Murre, C., McCraw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., Cabrera, C. V., Buskin, J. N., Hauschka, S. T., Lassar, A. B., Weintraub, H. & Baltimore, D. (1989) *Cell* **58**, 537–544.
6. Jones, N. (1990) *Cell* **61**, 9–11.
7. Olson, E. (1990) *Genes Dev.* **4**, 1454–1461.
8. Weintraub, H., Davis, R., Tapscott, S., Thayer, M., Krause, M., Benezra, R., Blackwell, T. K., Turner, D., Rupp, R., Hollenberg, S., Zhuang, Y. & Lassar, A. (1991) *Science* **251**, 761–766.
9. O'Shea, E. K., Rutkowski, R. & Kim, P. S. (1989) *Science* **243**, 538–542.
10. Oas, T. G., McIntosh, L. P., O'Shea, E. K., Dahlquist, F. W. & Kim, P. S. (1990) *Biochemistry* **29**, 2891–2894.
11. Nakabeppu, Y., Ryder, K. & Nathans, D. (1988) *Cell* **55**, 907–915.
12. Halazonetis, T. D., Georgopoulos, K., Greenberg, M. E. & Leder, P. (1988) *Cell* **55**, 917–924.
13. Chiu, R., Boyle, W. J., Meek, J., Smeal, T., Hunter, T. & Karin, M. (1988) *Cell* **54**, 541–552.
14. Kouzarides, T. & Ziff, E. (1989) *Nature (London)* **340**, 568–571.
15. Turner, R. & Tjian, R. (1989) *Science* **243**, 1688–1694.
16. Smeal, T., Angel, P., Meek, J. & Karin, M. (1989) *Genes Dev.* **3**, 2091–2100.
17. Gentz, R., Rauscher, F. J., Abate, C. & Curran, T. (1989) *Science* **243**, 1695–1699.
18. O'Neil, K. T., Hoess, R. H. & DeGrado, W. F. (1990) *Science* **249**, 774–778.
19. Talanian, R. V., McKnight, C. J. & Kim, P. S. (1990) *Science* **249**, 769–771.
20. Angel, P., Hattori, K., Smeal, T. & Karin, M. (1988) *Cell* **55**, 875–885.
21. Angel, P., Smeal, T., Meek, J. & Karin, M. (1989) *New Biol.* **1**, 35–43.
22. Schüle, R., Rangarajan, P., Kliewer, S., Ransone, L. J., Bolado, J., Yang, N., Verma, I. M. & Evans, R. M. (1990) *Cell* **62**, 1217–1226.
23. Studier, F. W., Rosenberg, A. H. & Dunn, J. J. (1991) *Methods Enzymol.* **185**, 60–89.
24. Bohmann, D. & Tjian, R. (1989) *Cell* **59**, 709–717.
25. Binetruy, B., Smeal, T. & Karin, M. (1991) *Nature (London)* **351**, 122–127.
26. Boyle, W. J., Smeal, T., Defize, L. H. K., Angel, P., Woodgett, J. R., Karin, M. & Hunter, T. (1991) *Cell* **64**, 573–584.
27. Fried, M. & Crothers, D. M. (1981) *Nucleic Acids Res.* **9**, 6505–6525.
28. Nehls, M. C., Rippe, R. A., Veloz, L. & Brenner, D. A. (1991) *Mol. Cell. Biol.* **11**, 4065–4073.
29. Buchman, C., Skroch, P., Dixon W., Tullius, T. D. & Karin, M. (1990) *Mol. Cell. Biol.* **10**, 4778–4787.
30. Karin, M. (1991) in *Molecular Aspects of Cellular Regulation*, eds. Cohen, P. & Foulkes, G. (Elsevier, Amsterdam), Vol. 6, pp. 143–161.
31. Vogt, P. & Bos, T. J. (1990) *Adv. Cancer Res.* **55**, 1–35.
32. Weiss, M. A., Ellenberger, T. E., Wobbe, C. R., Harrison, S. C. & Struhl, K. (1990) *Nature (London)* **347**, 575–578.
33. Patel, L., Abate, C. & Curran, T. (1990) *Nature (London)* **347**, 572–574.
34. Vinson, C. R., Sigler, P. B. & McKnight, S. L. (1989) *Science* **246**, 911–916.
35. Ryseck, R. F. & Bravo, R. (1991) *Oncogene* **6**, 533–542.
36. Kerppola, T. K. & Curran, T. (1991) *Cell* **66**, 317–326.
37. Clarke, N. D., Beamer, L. T., Goldberg, H. R., Berkower, C. & Pabo, C. (1991) *Science* **254**, 267–270.

Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 4892–4896, June 1993
Biochemistry

Design superiority of palindromic DNA sites for site-specific recognition of proteins: Tests using protein stitchery

CHANGMOON PARK*†, JUDY L. CAMPBELL†‡, AND WILLIAM A. GODDARD III*†§

*Materials and Molecular Simulation Center, Beckman Institute (139-74), †Division of Chemistry and Chemical Engineering, and ‡Division of Biology, California Institute of Technology, Pasadena, CA 91125

Contributed by William A. Goddard III, February 12, 1993

ABSTRACT **Using protein stitchery with appropriate attachment of cysteines linking to either C or N termini of the basic region of the v-Jun leucine zipper gene-regulatory protein, we constructed three dimers—pCC, pCN, and pNN. All three bind specifically to the appropriately rearranged DNA recognition sites for v-Jun: ATGAcgTCAT, ATGAcgATGA, and TCATcgTCAT, respectively (K_d, ≈4 nM at 4°C). Results of DNase I footprinting provide strong support for bent recognition helices in leucine zipper protein–DNA complexes. Comparison of the results for pCC and pNN with those for pCN shows the design superiority of palindromic sequences for protein recognition.**

The mechanism by which cells respond to external stimuli is a fundamental problem in modern biology. Transcriptional regulatory proteins are known to play a key role in several systems evolved by cells to convert extracellular signals into altered gene expression (1). They operate by specifically binding to DNA target sites, which regulate the transcription of particular genes. Prominent among transcriptional regulatory proteins are the leucine zipper family of proteins, which recognize the DNA binding site as either homodimers or heterodimers (2–4).

The leucine zipper proteins are characterized by two functional segments: (*i*) the leucine zipper region, a helical region containing four or five leucines spaced at seven-amino acid intervals, and (*ii*) the basic region containing many basic residues (5–10). The basic region appears to be unfolded in solution but assumes an α-helical structure binding to its recognition site (11–13). Site-directed mutagenesis (6, 7) and domain swapping (8–10) experiments show that the leucine zipper region mediates dimerization and that the basic region is responsible for DNA binding. Experiments replacing the leucine zipper region with a three-peptide linker (14, 15) showed that the dimerized basic region recognizes the same site as the native protein, supporting the scissors grip model (5), where each monomer recognizes the half site of the symmetrical DNA binding site. Recently, we showed that the normal dimer (denoted pCC), which selectively recognizes the sequence ATGAcgTCAT, can be inverted to form a protein (denoted pNN) that selectively recognizes the inverted site, TCATcgATGA (15).

Gel electrophoresis experiments (22) with Jun homodimer and with Jun–Fos heterodimer showed that Jun and Fos induce DNA bending in the opposite direction upon binding to the specific site. To explain this, it was proposed that the basic region of jun has a bent α-helix, while the basic region of Fos has a straight helix. However, a recent x-ray crystal structure (21) for the complex between GCN4 and DNA containing the GRE site (ATGACTCAT) showed a straight single α-helix for the basic region of GCN4. Our current results support the interpretation that the v-Jun homodimer bound to its specific site has bent α-helices.

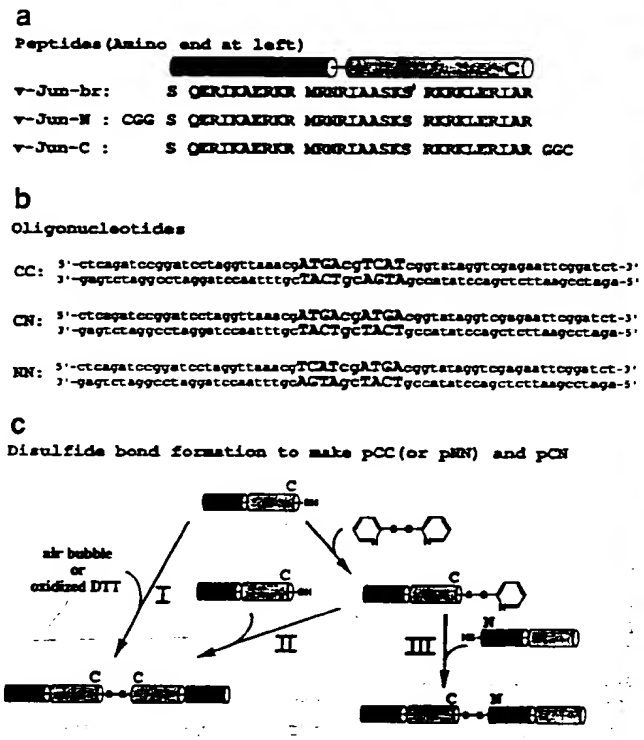

FIG. 1. Sequences of protein (*a*) and oligonucleotides (*b*) used in gel-retardation and footprinting studies. Total length of each oligonucleotide is 62. v-Jun-br contains the basic region of v-Jun. CGG is added to the N terminus of v-Jun-br to make v-Jun-N and GGC is added to the C terminus of v-Jun-br to make v-Jun-C. Proteins were chemically synthesized and checked by mass spectroscopy at the Biopolymer Synthesis Center at the California Institute of Technology (15). (*c*) Strategy for making pCC (and pNN) and pCN dimers. v-Jun-C was incubated with 10 mM dithiothreitol (DTT) for 5 hr at room temperature and purified directly into 10 mM 2,2′-dithiodipyridine/100 mM sodium phosphate, pH 5.5, containing 30% acetonitrile. Resulting thiopyridyl-(v-Jun-C) was purified by HPLC. Purified monomer v-Jun-N underwent reaction with 2 equivalents of thiopyridyl-(v-Jun-C) in solution containing 100 mM tetraethylammonium acetate buffer (pH 7.5) and 15% acetonitrile for 12 hr at room temperature. The final product, pCN, was purified by HPLC (15).

Peptide Design

Using protein stitchery, we have made three kinds of v-Jun (16, 17) homodimers (denoted pCC, pNN, and pCN) and show here that each selectively recognizes the appropriately reorganized DNA binding sites ATGAcgTCAT, TCATcgATGA, and ATGAcgATGA (see Fig. 1). The concept of protein stitchery (15) is that the individual basic arms (half

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

§To whom reprint requests should be addressed at *.

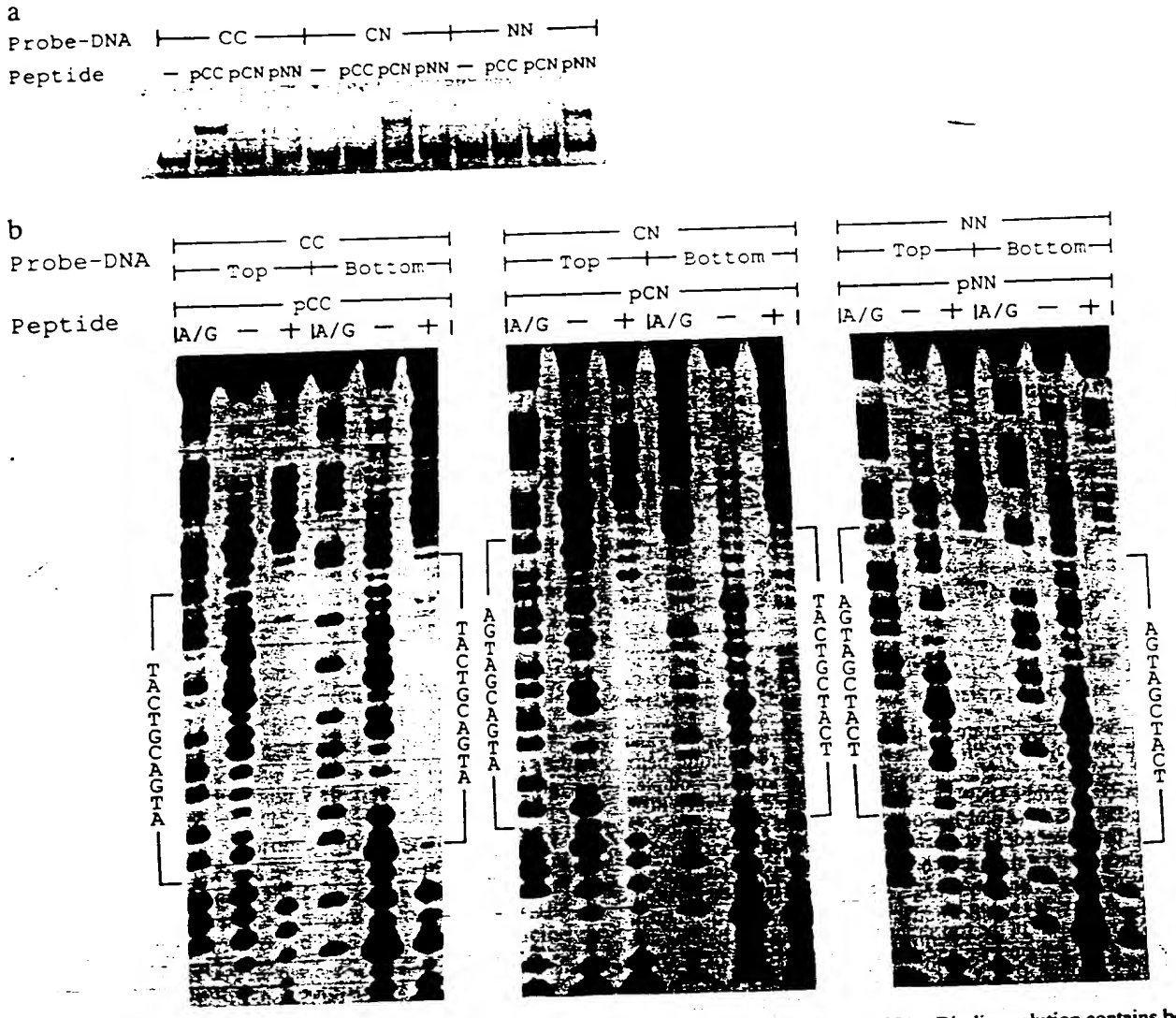

FIG. 2. (*a*) Gel-retardation assay for binding of pCC, pCN, and pNN to the CC, CN, and NN probe DNAs. Binding solution contains bovine serum albumin at 50 μg/ml, 10% (vol/vol) glycerol, 20 mM Tris·HCl (pH 7.5), 4 mM KCl, 4 mM $MgCl_2$, and the appropriate peptide at 3 nM in a 10-μl reaction volume. After 5000 cpm of each 5′ ^{32}P-labeled probe DNA was added, the solutions were stored at 4°C for 40 min and loaded directly on an 8% nondenaturing polyacrylamide gel in Tris/EDTA buffer at 4°C. As determined by titration of the gel shift, $K_d \approx 2$ nM for pCC/CC, $K_d \approx 6$ nM for pCN/CN, and $K_d \approx 4$ nM for pNN/NN, all at 4°C. These results show that each peptide binds specifically to its proposed binding site and not to the other sites. (*b*) DNase I footprinting assay of pCC, pCN, and pNN peptide with DNA containing the predicted binding sites for pCC, pCN, and pNN, respectively. Footprinting assay solution (in 50 μl) contains bovine serum albumin at 50 μg/ml, 5% (vol/vol) glycerol, 20 mM Tris·HCl (pH 7.5), 4 mM KCl, 2 mM $MgCl_2$, 1 mM $CaCl_2$, 20,000 cpm of each 5′ ^{32}P-labeled probe DNA, and the appropriate peptide at 50 nM. This solution was stored at 4°C for 40 min. After 5 μl of DNase I diluted in 1× footprinting assay buffer was added, the solutions were stored 1 min more at 4°C. DNase I digestion was stopped by addition of 100 μl of DNase I stop solution containing 15 mM EDTA (pH 8.0), 100 mM NaCl, and sonicated salmon sperm DNA at 40 μg/ml. This mixture was phenol/chloroform-extracted, ethanol-precipitated, and washed with 70% (vol/vol) ethanol. The pellet was resuspended in 5 μl of formamide loading buffer, denatured at 90°C for 4 min, and analyzed on 10% polyacrylamide sequencing gel (50% urea). These results show that each peptide specifically binds to the proposed binding site and protects the whole site except for the case of pCN/CN, which shows some incomplete protection on the binding site. This exception is explained as due to binding to semispecific (half) sites by single arms as discussed in the text (see Fig. 4).

sites) of the dimer and the individual half sites of the DNA can be recombined or stitched together in various sequences to form new proteins selective for binding to the new DNA sites. Thus, we use here the recognition helix v-Jun-br of Fig. 1*a* with a cysteine linker at either the N (v-Jun-N) or the C (v-Jun-C) terminus. These can be combined to form either pNN, pCC, or pCN dimers as illustrated in Fig. 1*c*. Formation of pNN and pCC (via pathway I) is straightforward since each involves dimerization of identical monomers. To ensure formation of pCN, the cysteine at the C terminus of v-Jun-C was reacted with excess 2,2′-dithiodipyridine to form thiopyridyl-(v-Jun-C) (18, 19) and then coupled with the cysteine at the N terminus of v-Jun-N to form the pCN dimer (v-Jun-C)-S–S-(v-Jun-N) (pathway III; Fig. 1*c*). We also verified pathway II for forming pCC.

Results and Discussion

We carried out gel-retardation assays (15) for each of the three peptide dimers with oligonucleotides (Fig. 1*b*) corresponding to each of the three proposed binding sites. These results (Fig. 2*a*) show that each dimer recognizes the appropriate binding site specifically with no detectable binding to the other sites. It is important to note that this strong preference for dimer occurs even though all oligonucleotides contain proper sites for binding a single arm of each dimer.

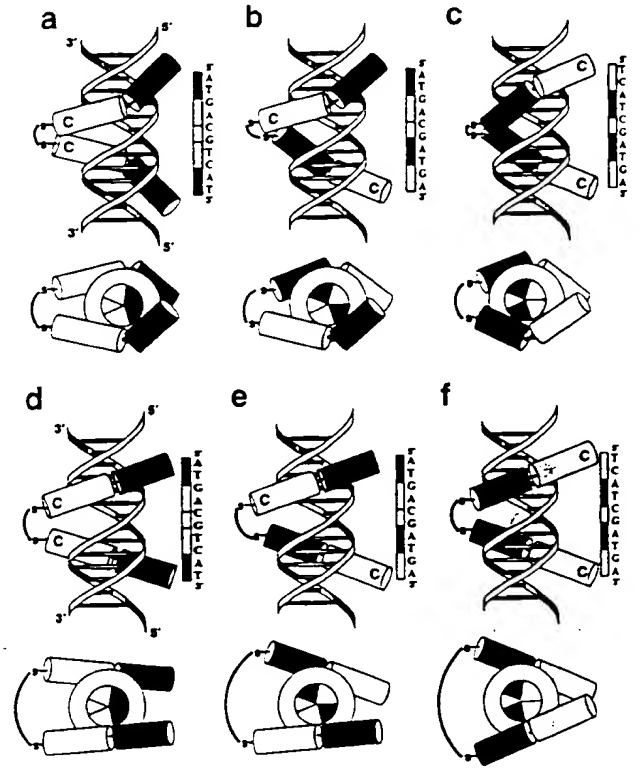

FIG. 3. Schematic diagram for the complex between peptides and their corresponding DNA sites assuming a bent recognition helix (*a–c*) and a straight recognition helix (*d–f*). (*a* and *d*) Complex between pCC and probe DNA CC. (*b* and *e*) Complex between pCN and probe DNA CN. (*c* and *f*) Complex between pNN and probe DNA NN. The linker connecting two monomers indicates a disulfide bond between the cysteines attached to the end of peptides. In each case, the side view is on the top and the top view is on the bottom. Outer and inner circles of the top view represent the outer and inner major groove surfaces of the top strand for the proposed binding site projected onto an imaginary plane perpendicular to the axis of DNA and running through the center of the peptide and binding site. Shading is used with the peptide and DNA contacts to ease the tracking of these regions in different cases. This diagram shows that a bent recognition helix can contact the same 4 bases for all three peptide dimers, while a linear recognition helix would contact different bases in the three peptide dimers. (This diagram is not meant to imply an exact correlation between where the basic region is bent and where the bases are positioned.)

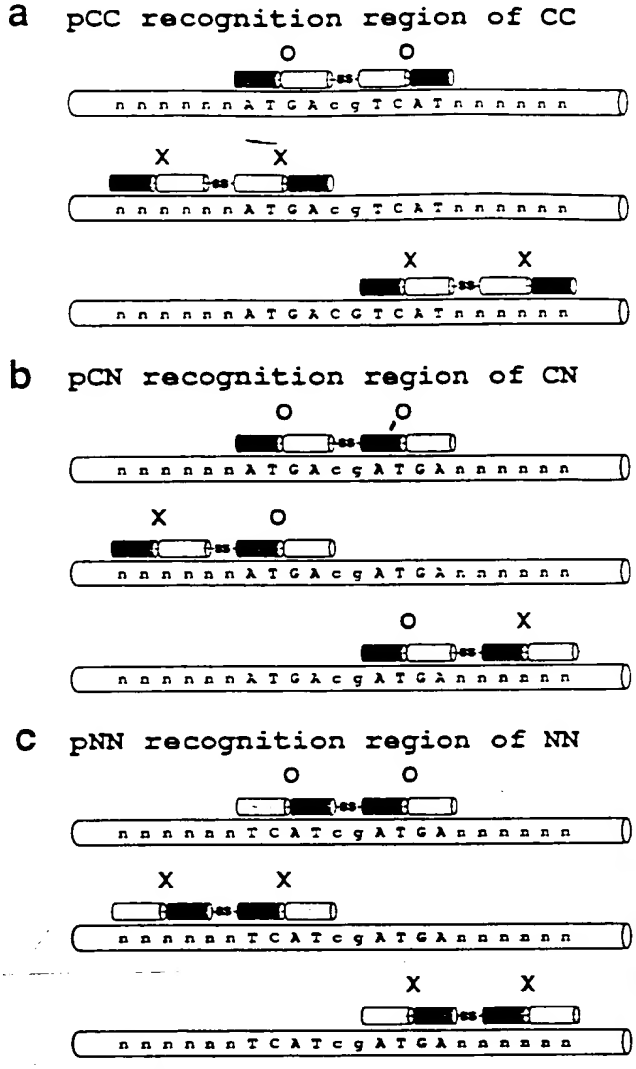

FIG. 4. Specific binding of protein at and near the corresponding DNA binding site. (*a*) Complex pCC/CC. (*b*) Complex pCN/CN. (*c*) Complex pNN/NN. O represents specific binding; X represents nonspecific binding. pCN/CN has one specific binding site and two sites for semispecific (half site) binding near its (nonpalindromic) binding site. However, pCC/CC and pNN/NN do not allow semispecific binding near their (palindromic) binding sites.

Therefore, at 3 nM peptide concentration the dimer does not make a stable complex with DNA unless both arms in the dimer recognize their proper sites. This implies cooperation between the monomers in recognizing the binding site (20). Since all three dimers have similar (2–6 nM) binding affinities with their own sites and since all three lead to the same length region protected from DNase I digestion (see below), we conclude that (*i*) all three cases involve similar conformations in the complex between DNA and peptide, and (*ii*) the monomer arm retains the same contact region in various dimers; this occurs despite the changing orientation of the monomers in the various peptide dimers (15).

There are two major models for the bound conformation of leucine zipper protein to the specific site. One is the induced helical fork model (13), which proposes a straight single α-helix for the basic region, and the other is a scissors grip model (5) which proposes a bent α-helix for the basic region. The recent x-ray crystal structure (21) for the complex of GCN4 containing only the basic and leucine zipper region and DNA-containing GRE site showed that the basic region of each protein has a straight α-helix conformation recognizing each half site of the dimer binding site. There was no DNA bending caused by protein binding (21). However, there remain many problems with assuming that the basic region is in all cases a straight α-helix: (*i*) The bases flanking the active site affect the binding of leucine zipper protein even though the crystal structure shows no direct contacts with protein (21). (*ii*) Gel electrophoresis experiments using Jun homodimer and Jun-Fos heterodimer showed that Jun and Fos induce DNA bending in opposite directions upon binding to their site (22), whereas GCN4 does not induce DNA bending (21, 27). (*iii*) Even though GCN4-br (a peptide containing the basic region of GCN4 protein) showed no specific binding (for details see ref. 14), we find that the monomer v-Jun-br (a peptide containing only the basic region of v-Jun; see Fig. 1*a*) specifically binds to the dimer site and shows the same protection as the dimer. Our conclusion then is that there is no universal model for the DNA-bound conformation of the basic region of leucine zipper proteins. Whether it is linear (as in GCN4) or bent (as in Jun) depends on the specific primary sequence and the properties of the solutions (stabilizers, pH, etc.) used in the experiments.

The result that all three dimers (pCC, pNN, and pCN) bind trongly to the appropriate combination of oligonucleotide ites implies that the helical binding arm is bent (5, 22) (see Fig.). Our argument is as follows. The optimum binding site for oth Jun homodimer and the Jun–Fos heterodimer is known to e ATGAcTCAT or ATGAcgTCAT, where the inner 7 or 8 ases play an important role in recognition (23, 24). The x-ray rystal structure of GCN4 bound to DNA leads to straight -helices, which have direct contacts with only the inner 7 ases of the GRE site (ATGACTCAT). Thus, each arm ecognizes the half-site (gATGAc or gTCATc) of the dimer inding site asymmetrically. If the same were true for v-Jun and if the same contacts are maintained between the protein and bases for the bound conformations of pCC/CC, pNN/NN, and pCN/CN (as expected since the binding constants and protection are the same), then the orientations of the binding arms would have very different orientations (Fig. 3 *e* and *f*). This should result in different protection from DNase I digestion (not observed). In addition, for the pNN/NN complex, this would lead to N termini of the two arms too distant to be connected by the added linker, GGCCGG. The alternative to Fig. 3 *d–f* is for each dimer to have the same angle (as in Fig. 3*d*). Thus, the actual contact region would not be equivalent in the three cases and it would be difficult to explain the gel retardation and footprinting results. Thus, we conclude that for v-Jun the basic region becomes bent upon binding to the DNA.

On the other hand, with the recognition helix bent roughly at the middle of the helix (as indicated in Fig. 3 *a–c*), it is plausible that the contact regions are ATGAcgTCAT for pCC, TCATcgATGA for pNN, and ATGAcgATGA for pCN. This leads to equivalent contact regions in all three cases and to the roughly equivalent binding energies apparent in Fig. 2*a*. In addition, footprinting (15) of the three peptide dimers (Fig. 2*b*), each with the appropriate oligonucleotide dimer, suggests that the complexed peptide dimers protect the full specific site (all 10 bp) from DNase I digestion. These results strongly support the bent recognition helix model for the basic dimers considered here and hence also for the leucine zipper parent proteins (21, 22).

For the pCN/CN complex, footprinting (Fig. 2*b*) shows incomplete protection on the binding site and partial protection on the bases flanking the binding site, whereas for pCC/CC and pNN/NN this does not happen. This occurs even though gel-retardation assays indicate specific binding for all complexes. Our explanation of this (Fig. 4) suggests why palindromic sequences are so common for selective binding of regulatory proteins (25, 26). This reasoning is supported by recent results we have observed showing that (*i*) the monomer of v-Jun containing only the basic region (v-Jun-br) specifically protects both pCC and pNN binding sites identically to the protection provided by the dimers pCC and pNN, respectively; (*ii*) at 3 nM concentration, gel retardation showed that pCC (and pCN) has lower binding affinity for the DNA probe carrying a sequence of cgATGAcgTCATcgTCATcg (containing pCC and pCN binding sites overlapping half of each dimer binding site in the center) than for CC (and CN) probe DNA. These results imply that the half site, gTCATc (or gATGAc), added next to the pCC (or pCN) binding site interferes with the binding of pCC (or pCN) to the dimer binding site (because the half site can be used as a binding site for each arm of the dimer if the orientation between the site and arm fits). Details of these results will be published elsewhere. Fig. 4 indicates the strength of binding for all three peptide dimers at or near their DNA recognition sites. Here, O represents good binding, while X represents nonspecific binding. The palindromic sites for pNN and pCC lead to binding only when the protein is exactly at the recognition site, whereas pCN can recognize both full site (both arms bound) and half sites (one arm bound). In gel retardation and DNase I footprinting, semispecific binding competes with specific binding. This occurs because one arm of the semispecifically bound peptide would cover half of the specific binding site, preventing another dimer from binding and providing full protection. This explains (*i*) why gel-retardation assays (Fig. 2*a*) show lower binding affinity for the pCN/CN complex compared to the pCC/CC and pNN/NN complexes and (*ii*) why footprinting assays (Fig. 2*b*) show incomplete protection on the binding site and partial protection on a few bases flanking the binding site. Such semispecific binding interferes with the site-specific binding and would eventually result in low production and abnormally slow growth. However, gel retardation shows no detectable nonspecific or semispecific binding at low peptide concentration, indicating that semispecific binding is significantly weaker than specific binding. After dimerization, the proteins suitable for palindromic dimer binding sites avoid semispecific DNA binding, leading to more selective recognition of the specific sites. Thus, palindromic dimer binding sites provide a good design for selective molecular recognition and for further flexibility the link can align sites (Fig. 3) to modify recognition.

The results on the three dimers considered here provide encouragement that this protein stitchery approach is feasible for designing and synthesizing proteins to recognize long DNA sequences. Thus, for trimers to recognize 15-bp sequences, we are using an approach similar to that of Fig. 1*c* involving appropriate use of cysteine linkages and transfer activators. It seems possible to design proteins for 20 bp and longer.

In summary, we find the following: (*i*) Protein stitchery of v-Jun leads to three dimers (pCC, pNN, and pCN), each of which binds specifically to the appropriate rearrangement of DNA sites. Thus, there is cooperation between the two monomers of the dimer in binding to DNA, which depends on the relative orientation of two monomers in the dimer. (*ii*) These results provide strong support for the bent α-helix model of the basic region when bound to DNA. (*iii*) These results provide an explanation for the advantage of dimerization and the use of palindromic sites in the site-selective binding of proteins to DNA. (*iv*) These results show protein stitchery to be useful for establishing the conformation and mechanism for binding of proteins to their DNA binding sites.

This is contribution no. 8776 from the Division of Chemistry and Chemical Engineering, California Institute of Technology. We thank Prof. Peter Kim for helpful criticisms. This work was supported by a grant from the Department of Energy-Advanced Industrial Concepts Division. The facilities of the Materials and Molecular Simulation Center are also supported by grants from the National Science Foundation-CHE 91-100289), the National Science Foundation-Advanced Computing, the Department of Energy-Chemical Biology, Allied Chemical, Asahi Chemical, Asahi Glass, Chevron, BF Goodrich, BP America, and Xerox.

1. Curran, T. & Franza, B. R., Jr. (1988) *Cell* 55, 395–397.
2. Kouzarides, T. & Ziff, E. (1989) *Nature (London)* 340, 568–571.
3. Turner R. & Tjian, R. (1989) *Science* 243, 1688–1694.
4. Gentz, R., Rauscher, F. J., Abate, C. & Curran, T. (1989) *Science* 243, 1695–1699.
5. Vinson, C. R., Sigler, P. B. & McKnight, S. L. (1989) *Science* 246, 911–916.
6. Neuberg, M., Schermann, M., Hunter, J. B. & Muller, R. (1989) *Nature (London)* 338, 589–590.
7. Ransone, L. J., Visvader, J., Wamsley, P. & Verma, I. M. (1990) *Proc. Natl. Acad. Sci. USA* 87, 3806–3810.
8. Ransone, L. J., Wamsley, P., Mosley, K. L. & Verma, I. M. (1990) *Mol. Cell. Biol.* 10, 4565–4573.
9. Agre, P., Johnson, P. K. & McKnight, S. L. (1989) *Science* 246, 922–926.
10. Neuberg, M., Adamkiewicz, J., Hunter, J. P. & Muller, R. (1989) *Nature (London)* 341, 243–245.
11. Weiss, M. A. (1990) *Biochemistry* 29, 8020–8024.

12. Patel, L., Abate, C. & Curran, T. (1990) *Nature (London)* **347**, 572–575.
13. O'Neil, K. T., Hoess, R. H. & DeGrado, W. F. (1990) *Science* **249**, 774–778.
14. Talanian, R. V., McKnight, C. J. & Kim, P. S. (1990) *Science* **249**, 769–771.
15. Park, C., Campbell, J. L. & Goddard, W. A., III (1992) *Proc. Natl. Acad. Sci. USA* **89**, 9094–9096.
16. Kaki, Y., Bos, T. J., Davis, C., Starbuck, M. & Vogt, P. K. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 2848–2852.
17. Bos, T. J., Bohmann, D., Tsuchie, H., Tjian, R. & Vogt, P. K. (1988) *Cell* **52**, 705–712.
18. Corey, D. & Schultz, P. G. (1987) *Science* **238**, 1401–1403.
19. Zuckermann, R., Corey, D. & Schultz, P. G. (1987) *Nucleic Acids Res.* **15**, 5305–5321.
20. Abate, C., Luk, D., Gagne, E., Roeder, R. G. & Curran, T. (1990) *Mol. Cell. Biol.* **10**, 5532–5535.
21. Ellenberg, T. E., Brandl, C. J., Struhl, K. & Harrison, S. C. (1992) *Cell* **71**, 1223–1237.
22. Kerpolla, T. K. & Curran, T. (1991) *Science* **254**, 1210–1214.
23. Ryseck, R. & Bravo, R. (1991) *Oncogene* **6**, 533–542.
24. Schule, R., Umesono, K., Mangelsdorf, D. J., Bolado, J., Pike, J. W. & Evans, R. M. (1990) *Cell* **61**, 497–504.
25. Pabo, C. O. & Sauer, R. T. (1984) *Annu. Rev. Biochem.* **53**, 293–321.
26. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988) *Science* **240**, 1759–1764.
27. Gartenberg, M. R., Ampe, C., Steitz, T. A. & Crothers, D. M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6034–6038.

Design and Synthesis of a New Peptide Recognizing a Specific 16-Base-Pair Site of DNA

Changmoon Park,[†,‡,#] Judy L. Campbell,[†,§] and William A. Goddard, III*[,†,⊥]

Contribution from the Materials and Process Simulation Center, Beckman Institute (139-74), Division of Chemistry and Chemical Engineering (CN 8920), and Division of Biology, California Institute of Technology, Pasadena, California 91125

Received January 28, 1994. Revised Manuscript Received July 18, 1994[⊗]

Abstract: We designed a peptide to recognize a new 16-base-pair site (about 1.5 turns) of DNA by stitching together three peptides of the v-Jun basic region in a specified order. The binding site consists of three five-base-pair half-sites each of which is recognized by a different segment of the peptide. DNase I footprinting shows that the new peptide specifically recognizes the proposed site, and gel retardation shows that the dissociation constant is about 5 nM at 4 °C. Gel retardation shows that the new peptide does recognize the proposed trimer binding site about 10 times stronger than the dimer binding sites [having two half-sites for two arms]. These results also provide information about the relationship between specific and nonspecific binding in the recognition between protein and DNA.

1. Introduction

Proteins that bind selectively to a specific DNA binding site play important roles in biological systems. Thus the regulation of cellular reactions (including replication, transcription, and translation) is mostly mediated by the specific interactions of DNA binding proteins with DNA.[1] As a result, design and synthesis of sequence-specific DNA binding proteins are of great interest in modern chemical biology.

Synthesis of peptides specifically recognizing long sequences (more than 10 base pairs (bp's)) of DNA is also important in mapping large genomes. Most known restriction enzymes recognize 4−8-bp sites, creating too many fragments to be handled when used to digest genomic DNA. Many attempts have been developed to recognize (and cleave) specific longer sites of DNA.[2−8] However, most of the current methods are indirect, requiring a series of steps (protection, chemical modification, and deprotection) to obtain the desired results.

We illustrate here the protein stitchery approach for designing a new protein to recognize a specific long site (16 bp's) of DNA. This is illustrated in Figure 1, which contains three fragments each corresponding to the basic region of v-Jun.

v-Jun is a member of the leucine zipper protein class of regulatory proteins for DNA transcription. It binds as a homodimer or as a heterodimer with Fos to a DNA site having dyad symmetry.[10,11] A recent X-ray crystal structure for the complex of GCN4 (another leucine zipper protein) and DNA[12,19] shows that the dimerization is mediated by the leucine zipper region and that each basic region forms an α-helix as it recognizes the half-site of the dimer binding site. The α-helix of the protein−DNA complex may bend depending on the nature of the binding site. In the absence of the specific DNA binding site, the basic region of the leucine zipper protein has a flexible structure in solution. However, it changes to α-helix when bound to the specific site of DNA.[22,23]

We use gel retardation and footprinting assays to show that the new peptide stitched together from three v-Jun basic regions

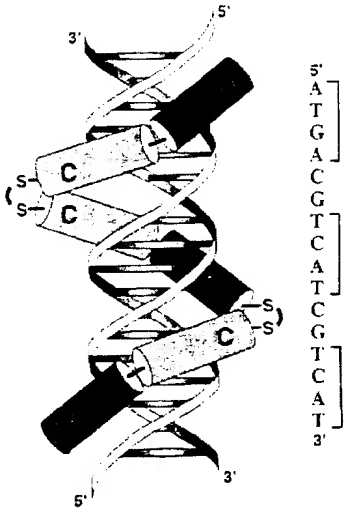

Figure 1. Schematic diagram for the complex between the peptide trimer pCC·NC and the trimer binding site o-CC·NC. The proposed binding site sequence for the top strand of DNA (see Figure 2b) is shown on the right of the diagram. The current and previous experimental results[13−15] suggest that the peptide wraps around the DNA along the major groove to recognize all three monomer binding sites.

[†] Beckman Institute.
[‡] Division of Chemistry and Chemical Engineering.
[§] Division of Biology.
[#] Current address: Chungnam National University, Department of Chemistry, Taejon, South Korea.
[⊥] To whom correspondence should be addressed.
[⊗] Abstract published in *Advance ACS Abstracts*, June 1, 1995.

(1) Pabo, C. O.; Sauer, R. T. *Annu. Rev. Biochem.* **1984**, *53*, 293−321.
(2) McClelland, M.; Kessler, L.; Bittner, M. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 983−987.
(3) Pei, D.; Corey, D. R.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9858−9862.
(4) Koob, M.; Grimes, E.; Szybalski, W. *Science* **1988**, *241*, 1084−1086.
(5) Strobel, S. A.; Dervan, P. B. *Science* **1990**, *249*, 73−75.
(6) Ferrin, L. J.; Camerini-Otero, R. D. *Science* **1991**, *254*, 1494−1497.
(7) Sluka, J. P.; Horvath, S. J.; Bruist, M. F.; Simon, M. I.; Dervan, P. B. *Science* **1987**, *238*, 1129−1132.
(8) Mack, D. P.; Iverson, B. L.; Dervan, P. B. *J. Am. Chem. Soc.* **1988**, *110*, 7572−7574. Strobel, S. A.; Dervan, P. B. *Science* **1990**, *249*, 73−75.
(9) Struhl, K. *Cell* **1987**, *50*, 841−846.
(10) Kaki, Y.; Bos, T. J.; Davis, C.; Starbuck, M.; Vogt, P. K. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 2848−2852.
(11) Bos, T. J.; Bohmann, D.; Tsuchie, H.; Tjian, R.; Vogt, P. K. *Cell* **1988**, *52*, 705−712.
(12) Ellenberg, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell* **1992**, *71*, 1223−1237.

0002-7863/95/1517-6287$09.00/0

EXHIBIT D

(a) Peptides (amino terminus on the left)

```
v-Jun-br:      S  QERIKAERKR  MRNRIAASKS  RKRKLERIAR
v-Jun-N : CGG  S  QERIKAERKR  MRNRIAASKS  RKRKLERIAR
v-Jun-C :      S  QERIKAERKR  MRNRIAASKS  RKRKLERIAR  GGC
v-Jun-NC: CGG  S  QERIKAERKR  MRNRIAASKS  RKRKLERIAR  GGC
```

(b) Oligonucleotides

```
o-CC•NC : 5'-ctcagatccggatcctaggttaaacgATGAcgTCATcgTCATcggtataggtcgagaattcggatcct-3'
          3'-gagtctaggcctaggatccaatttgcTACTgcAGTAgcAGTAgccatatccagctcttaagcctagga-5'

o-CC    : 5'-ctcagatccggatcctaggttaaacgATGAcgTCATcggtataggtcgagaattcggatcct-3'
          3'-gagtctaggcctaggatccaatttgcTACTgcAGTAgccatatccagctcttaagcctagga-5'

o-CN    : 5'-ctcagatccggatcctaggttaaacgATGAcgATGAcggtataggtcgagaattcggatcct-3'
          3'-gagtctaggcctaggatccaatttgcTACTgcTACTgccatatccagctcttaagcctagga-5'
```

(c) Procedure to make pCC•NC

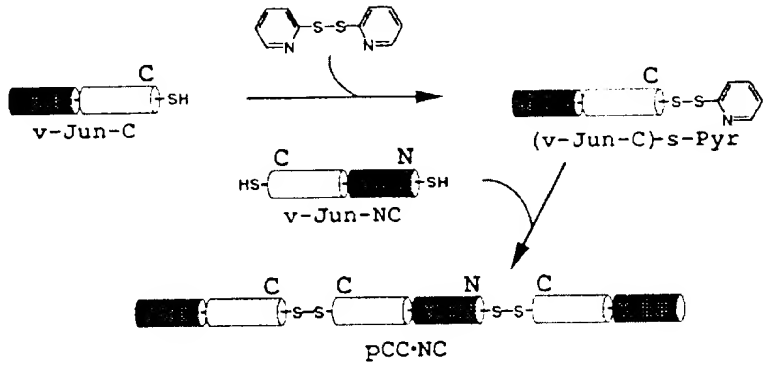

Figure 2. Sequences of protein (a) and oligonucleotides (b) used in the gel retardation and footprinting studies. The total length of oligonucleotides o-CC, o-CN, and o-CC•NC are 62, 62, and 68, respectively. The peptide v-Jun-br contains the basic region of v-Jun (amino acids 214–244).[7] v-Jun-N and v-Jun-C were prepared as described previously.[13,14] v-Jun CN (which is equivalent to v-JUN-NC) was chemically synthesized and purified, and the purity was checked by mass spectroscopy at the Biopolymer Synthesis Center at the California Institute of Technology as described previously.[13,14] calcd, 4256.4; exptl, 4259.1. (c) Strategy for making the pCC•NC trimer.

binds selectively to the 16-bp site of DNA composed of three half-sites (appropriately oriented) for the v-Jun dimer. These results provide further insight on the interaction of leucine zipper proteins with DNA.

2. Materials and Experiments

2.1. Peptides and Oligonucleotide Synthesis. The peptide monomers v-Jun-N, v-Jun-C, and v-Jun-CN (see Figure 2a) were prepared as described previously.[13,14] The automated stepwise syntheses were done on an Applied Biosystems model 430A peptide synthesizer with an optimized synthetic protocol of the *N-tert*-butoxycarbonyl (t-Boc) chemistry. The peptides were purified by reverse-phase high-performance liquid chromatography (HPLC) on a Vydac C18 column. A linear gradient of 0–50% aqueous/acetonitrile/0.1% trifluoroacetic acid was run over 120 min.

The oligonucleotides o-CC, o-CN, and o-CC•NC (indicating an oligonucleotide containing the proposed binding site of pCC•NC; see below for the notation) were used to mimic various DNA binding sites as shown in Figure 2b. These were synthesized using the facilities at the Biopolymer Synthesis Center at Caltech. o-CC has the binding site (ATGAcgTCAT) of the v-Jun dimer while the others are formed with various rearrangements of the half-site. The synthesized oligonucleotides were purified by using 10% denaturing polyacrylamide gel, and duplexes were made between complementary oligonucleotides if needed.

2.2. Synthesis of the Peptide Dimer and Peptide Trimer. The procedure to synthesize homodimer pCC is straightforward. In oxidizing conditions (5 mM oxidized dithiothreitol) v-Jun-C dimerizes to form pCC. However, to synthesize heterodimer pCN requires additional steps. In order to form pCN without also forming pCC and pNN, we activated the thiol group of v-Jun-C using 2,2′-dithiodipyridine (see Figure 2c) and purified the resulting thiopyridyl-(v-Jun-C) with HPLC.[14,15] This was reacted with purified v-Jun-N to form pCN.

To form the trimer pCC•NC (indicating a peptide trimer consisting of three monomer arms connected by two disulfide bonds; one is made between two C-termini of the first and second arms, and the other one is made between the N-terminus of the second arm and the C-terminus of the third arm), we used a similar procedure in which purified monomer v-Jun-NC was reacted with excess (3 equiv) thiopyridyl-(v-Jun-C) to make the trimer product pCC•NC (see Figure 2c), which was purified by HPLC.[14,15] To verify the formation of peptide heterodimer and peptide heterotrimer, HPLC analyses were done with the purified pCC•NC and pCN and with pCC•NC and pCN reduced by 20 mM dithiothreitol (DTT). The HPLC analysis (Figure 3) showed that reduction of pCC•NC yields only the two peaks corresponding to v-Jun-NC and v-Jun-C in the expected 1:2 ratio, while pCN shows two peaks corresponding to v-Jun-C and v-Jun-N in the expected 1:1 ratio. This HPLC analysis confirms the formation of heterotrimer pCC•NC and heterodimer pCN because each of v-Jun-C and v-Jun-N has only one thiol group on one terminus, while v-Jun-CN has two thiol groups on

(13) Park, C.; Campbell, J. L.; Goddard, W. A., III. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9084–9096.
(14) Park, C.; Campbell, J. L.; Goddard, W. A., III. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 4892–4896.
(15) Park, C. Ph.D. Thesis, Caltech, May 1993.
(16) Corey, D.; Schultz, P. G. *Science* **1987**, *238*, 1401–1403. Zuckermann, R.; Corey, D.; Schultz, P. G. *J. Am. Chem. Soc.* **1988**, *110*, 1614–1615.
(17) Zuckermann, R.; Corey, D.; Schultz, P. G. *Nucleic Acids Res.* **1987**, *15*, 5305–5321.

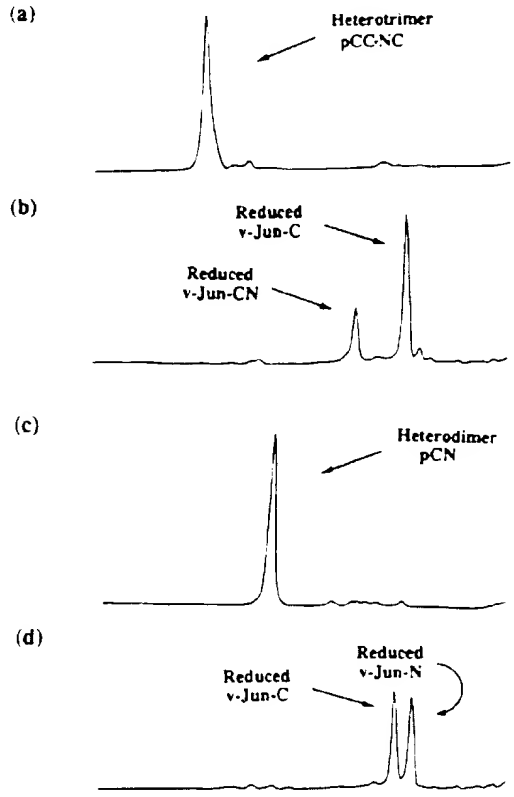

Figure 3. HPLC analysis of peptide dimer and trimer. (a) Purified heterotrimer pCC·NC and (b) reduced pCC·NC with 20 mM DTT for 4 h at 25 °C. (c) Purified heterodimer pCN and (d) reduced pCN as in b. The results of HPLC analysis show that the heterotrimer pCC·NC consists of 1 equiv of v-Jun-CN and 2 equiv of v-Jun C, while pCN consists of 1 equiv of v-Jun-C and 1 equiv of v-Jun-N as expected.

both termini, each of which is eligible to make a disulfide bond with another thiol group.

2.3. Gel Retardation and Footprinting Assays. Gel retardation and footprinting assays were carried out as described previously.[13c,5] The binding solution of gel retardation contains bovine serum albumin at 50 mg/mL, 10% (v/v) glycerol, 20 mM Tris-HCl (pH 7.5), 4 mM KCl, 2 mM $MgCl_2$, and 3 nM of appropriate peptides in 10 μL reaction volume. After adding 5000 cpm of each 5′-^{32}P-labeled probe DNA as indicated, the solutions were stored at 4 °C for 1 h and loaded directly on an 8% nondenaturing polyacrylamide gel in TE buffer at 4 °C. The gel was equilibrated for 2 h at 20 mA before the samples were loaded, and electrophoreses were performed for 3 h at 100 V at 4 °C after the samples were loaded.

The gel was dried and exposed to Kodak storage phosphor screen SO 230 (from Molecular Dynamics) in the dark room for 2 h. A Molecular Dynamics 400S PhosphoImager and IMAGEQUANT version 3.0 were used to integrate the volume of each rectangle drawn around the free and bound bands in the same dimension (see Table 1).

The footprinting assay solution (in 50 L) contains bovine serum albumin at 50 mg/mL, 5% glycerol, 20 mM Tris-HCl (pH 7.5), 4 mM

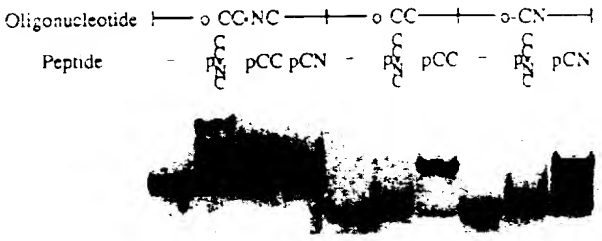

Figure 4. Gel retardation assays for binding of pCC·NC, pCC, and pCN to o-CC·NC, o-CC, and o-CN. These studies were carried out as described in the text. A 3 nM solution of each peptide was used in a 10 μL reaction volume containing 5000 cpm of the appropriate oligonucleotide.

KCl, 2 mM $MgCl_2$, 1 mM $CaCl_2$, and 20 000 cpm of each 5′-^{32}P-labeled probe DNA (60−62 bp) and 50 nM v-Jun-NN. This solution was stored at 4 °C for 1 h. After adding 5 μL of DNase I diluted in 1× footprinting assay buffer, the solutions were stored 1 min more at 4 °C. The DNase I digestion was stopped by addition of 100 μL of DNase I stop solution containing 15 mM EDTA (pH 8.0), 100 mM NaCl, 25 μg/mL sonicated salmon sperm DNA, and 25 μg/mL yeast tRNA. This was phenol/chloroform extracted, ethanol precipitated, and washed with 70% ethanol. The pallet was resuspended in 5 μL of formamide loading buffer, denatured at 90 °C for 4 min, and analyzed on 10% denaturing polyacrylamide sequencing gel (50% urea).

3. Results

3.1. Specificity of pCC·NC for o-CC·NC. The gel retardation assays (Figure 4) show that pCC·NC binds to o-CC·NC, which has the exact site designed to simultaneously bind all three arms of pCC·NC. On the basis of gel shift titrations the binding constant is about 5 nM (see Table 1). However, the gel retardation assays show very weak binding (40−50 nM) of pCC·NC to o-CC or o-CN, each of which has a site for two arms of pCC·NC. Combined with the results for v-Jun homodimers, this indicates that pCC·NC makes contact with about 16 bp's of DNA (about 1.5 turns of duplex DNA) along the major groove (see Figure 1).

The DNase I footprinting assays (Figure 5) show that the new peptide pCC·NC protects the full proposed binding site, confirming the results from gel retardation assays. These results indicate that each of the three arms of pCC·NC binds to the proposed half-site, protecting each of the three half-sites from DNase I digestion (see Figure 1).

3.2. Binding of pCC and pCN to the Dimer and Trimer Binding Site. The results of gel retardation show that pCC and pCN bind to their proposed binding site with dissociation constants of about 2 and 6 nM, respectively. This in good agreement with our previous experiments.[14] pCC and pCN bind to the trimer binding site, o-CC·NC, about three times more weakly than to the dimer binding sites, o-CC and o-CN, respectively. This indicates that the additional monomer binding site in o-CC·NC compared to the dimer binding site interferes with the dimers in binding to their dimer binding sites. This

Table 1. Results of Titration of the Gel Shift Using a Molecular Dynamics 400S PhosphorImager

	o-CC·NC				o-CC			o-CN		
peptide	no	pCC·NC	pCC	pCN	no	pCC·NC	pCC	no	pCC·NC	pCN
bound	3682[a]	35469	35522	12434	2479[a]	5802	38385	3056[a]	10243	46431
free	72854	47297	52981	61928	61515	58087	20701	89452	88586	86518
ratio[b]		0.645	0.601	0.141		0.057	1.735		0.081	0.501
K_d[c] (nM)		4.7	5.0	21.3		52.6	1.7		37.0	6.0
ΔG_s[d] (kcal/mol)		10.6	10.5	9.7		9.2	11.1		9.4	10.4

[a] These values are used to correct the background for the bound band, used as (background) below. [b] Ratio = (bound)*/(free), where (bound)* = (bound) − (background) as described in a. [c] $K_d = 1/K_a = [P][D]/[PD] = [P](free)/(bound)$, where [P],[D], and [PD] indicate the concentrations of peptide, DNA binding site, and peptide/DNA complex, respectively. [d] $\Delta G_s = -RT \ln K_a = RT \ln K_d$ at $T = 277.15$ K (4 °C).

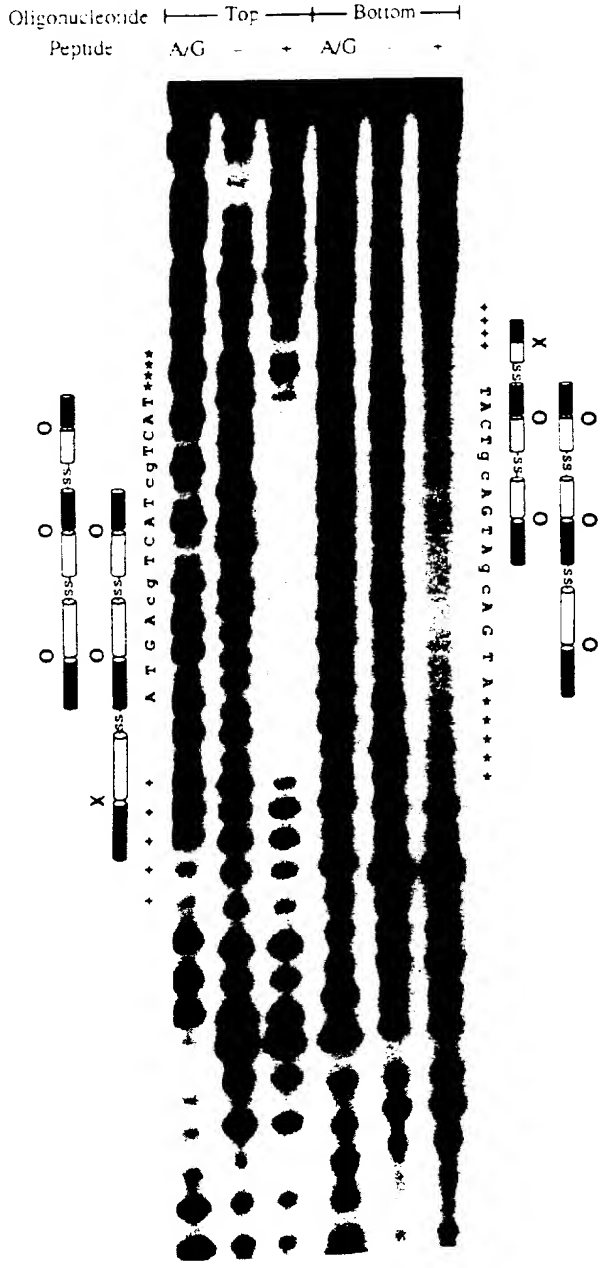

Figure 5. DNase I footprinting assays of pCC·NC with o-CC·NC were performed as described in the text. A 50 nM solution of pCC·NC was used with 50 000 cpm of o-CC·NC in a 50 μL reaction volume. The first column to the left and right shows the sequence of the o-CC·NC active site, and the outer column on each side shows the p-CC·NC protein bonded to this site. Clearly this entire region is protected. In addition the two sites next to the active site will generally show protection. However the observations show additional protection on the 5′ side of the top strand and the 5′ side of the bottom strand. Additional protection is given next to the last column on each side which shows the semispecific binding of the pCC·NC protein to the o-CC site. This leads to exactly the additional protection of the sites labeled as ++−− (5′ side of site) but not to protection of the sites labeled **** on the 3′ side.

implies that there might be some direct interaction between the dimer and the added monomer binding site or that the added monomer binding site affects the binding of dimer to the neighboring dimer binding site indirectly in an unknown way (for example, by changing the conformation of DNA).

3.3. Binding of Heterotrimer to Dimer Binding Site. The heterotrimer binds specifically to the proposed trimer binding site. However, the gel retardation results show that there is also a weak binding to the dimer binding sites. Gel titration (Table 1) shows that the heterotrimer pCC·NC binds to the proposed trimer binding site, o-CC·NC, about 10 times more strongly than to the dimer binding sites, o-CC and o-CN. This is equivalent to a free energy difference of about 1.3 kcal/mol. In another words, the third arm of the heterotrimer stabilizes the trimer by about 1.3 kcal/mol when bound to the trimer binding sites compared to the dimer binding site. However, for binding to o-CC, the third arm destabilizes the binding relative to pCC by 1.9 kcal/mol and, for binding to o-CN, the third arm destabilizes the binding relative to pCN by 1.0 kcal/mol. This results in destabilizing the binding of the other two monomers to the dimer binding site by about 30 times for o-CC and about six times for o-CN. Therefore, compared to the dimers, the additional arm of pCC·NC trimer destabilizes the binding of the trimer to the imperfect binding site while it stabilizes the binding of trimer to the trimer binding site.

3.4. Semispecific Binding of pCC·NC. The footprinting studies also provide some evidence for *semispecific binding* in which the pCC·CN protein is reversed so that it recognizes only the o-CC binding site of o-CC·NC. The location of the p-CC·CN protein on the full o-CC·NC binding site is indicated by the outer columns of Figure 5 (where O indicates specific binding). The reversed p-CC·NC protein can also recognize the o-CC region as indicated in the next to the last column of Figure 5 (here X indicates nonspecific binding), which is about 10 times weaker than the specific binding.

In such semispecific binding the nonspecifically bound arm would create partial protection on the bases beyond the 5′ end of the protein site for the top strand and on the bases beyond the 3′ end of the protein binding site for the bottom strand. At the same time the semispecific binding would lead to incomplete protection on the 3′ end of the protein binding site on the top strand and of the 5′ end on the bottom strand. Therefore, such semispecific binding would result in a quite asymmetric protection pattern around the binding site.

The results of DNase I footprinting (Figure 5) show this expected asymmetry. For the top strand of DNA, the partially protected region is expanded far beyond the protein binding site (up to the seventh base) in the 5′ region, whereas for bottom strand of DNA, the last two base pairs in the 5′ region of protein binding site are not completely protected. The reverse situation occurs for the 3′ regions, where extra protection occurs for the bottom strand and less occurs for the top strand.

4. Discussion

Polypeptides can recognize more than one turn of DNA (that is, more than 10 bp's of DNA) in two ways: (1) by wrapping around the DNA along the major groove and (2) by approaching the binding site from one face of the DNA. Case 2 requires the polypeptide to also interact with the minor groove of DNA, while case 1 allows binding to only the major groove. Case 1 is much easier to design than case 2 because an α-helix fits nicely into the major groove of DNA but not into the minor groove. However, to wrap around the DNA, the polypeptide must be sufficiently flexible to follow the major groove of DNA along its helical pathway. If the structure of the polypeptide is too rigid, it cannot wrap around the DNA to recognize an additional turn of the DNA. The basic region of the leucine zipper protein is an ideal candidate to satisfy all these criteria. It has no fixed structure in solution in the absence of its specific DNA binding site, but it changes into an α-helix when bound to the specific DNA binding site. From our previous experi-

ments,[14] each one of the v-Jun basic regions exactly recognizes its monomer binding site independently of the relative orientation of the additional basic regions (connected through a disulfide bond between the thiol groups of cysteines added on the terminus of the peptide monomer).

The new results show that the new peptide trimer pCC·NC specifically binds to the proposed trimer binding site of o-CC·NC (see Figure 1) but also binds about 10 times more weakly to the dimer binding sites, o-CC or o-CN. This protein stitchery strategy can be used to design other new peptides for recognizing new or longer sites. Thus we would decompose the target site in terms of segments (three to five base pairs) each of which is recognized by a portion of a DNA binding protein. The DNA binding regions would then be stitched together to form the full protein for selectively recognizing the new site.

In order to measure accurate free energy differences for a peptide to a different DNA binding site, direct competition assays between the DNA binding sites are required. However, we can estimate the free energy difference (see Table 1) from the free energies calculated using the intensity of the bound and free bands in Figure 4. Our current results show that the third peptide arm of pCC·NC compared to the dimer (pCC or pCN) (1) stabilizes the binding of pCC·NC when it finds a perfect trimer binding site, o-CC·NC, and (2) destabilizes the binding of pCC·NC to the incomplete binding site (o-CC or o-CN), as compared with the dimers binding to the dimer binding sites. This provides an explanation for the results of our previous experiments[13-15] where each peptide dimer (pCC, pCN, and pNN) selectively recognized the proposed dimer binding sites (o-CC, o-CN, and o-NN, respectively) but not the binding sites selectively recognized by the other peptide dimers.

These studies provide additional observations that should be useful in elucidating the details of protein-DNA recognition. Thus pCC shows a binding affinity for the o-CC·NC site of about one-third of the affinity for o-CC even though o-CC·NC contains a binding site for pCC (Table 1). Similarly pCN shows a binding affinity[19] for the o-CC·NC site of about one-fourth of the affinity for o-CN even though o-CC·NC has a binding site for pCN. This implies that the half-site added next to the binding site of pCC (or pCN) to make the binding site of pCC·NC interferes with pCC (or pNN) in binding to the dimer binding site. Additional recent results[18] show that the basic region of v-Jun by itself recognizes the dimer binding site specifically without dimerization. This implies that the interaction between the monomer of basic region of v-Jun and the monomer binding site is strong enough to retain the complex. Therefore, it is reasonable to propose that a direct interaction between the dimer and the added monomer binding site in o-CC·NC compared to the o-CC (or o-CN) interferes with the dimer in binding to the neighboring dimer binding site. However, it may also be that other indirect effects interfere with the dimer binding site.

For the top strand of o-CC·NC there is partial protection on the 3′ end bases flanking the binding site. The reason for this partial protection is that p-CC·NC also exhibits semispecific binding to the o-CC portion of the site. Such semispecific binding is supported by the observation that the glucocorticoid receptor recognizes the incorrect spaced binding site semispecifically, with one subunit binding specifically with the correct half-site and the other nonspecifically with a noncognate site.[20] Similarly the basic region of GCN4 shows a relatively strong binding affinity for the randomized sequence of DNA,[21] indicating it is possible for the basic region to have a nonspecific interaction with the nonspecific sequence of DNA. Our results do not indicate if the affinity of nonspecific binding depends on the DNA sequence.

Comparing the binding of pCC·NC to o-CC·NC and pCC to o-CC, there is no gain in binding energy from dimer to trimer even though the trimer binds to the trimer binding site 10 times stronger than to the dimer binding site. These results suggest that the added linker on the terminus of the peptide monomer to replace the leucine zipper region is not flexible enough (or long enough) to wrap around 1.5 turns of DNA, resulting in strain on the trimer. This does not happen in the case of dimers because they need only to wrap around about one turn of DNA. Therefore a more flexible (or longer) linker than the present one (Gly-Gly-Cys) may improve the binding affinity of the trimer to the trimer binding site.

We are now in the process of using molecular modeling, molecular dynamics, and thermodynamic perturbation theory to determine the details concerning the protein DNA recognition and to explain the origins of the above results.

Acknowledgment. All peptides and oligonucleotides were synthesized using the facilities at the Biopolymer Synthesis Center at Caltech. This research was supported by a grant from the Biological and Chemical Technologies Research (BCTR) of the Department of Energy (DOE). The facilities of the Materials and Process Simulation Center (MSC) are also supported by grants from the National Science Foundation (CHE-91100284 and ASC-9217368), Allied Chemical, Asahi Chemical, Asahi Glass, Chevron Petroleum Technology Co., BF Goodrich, BP America, Teijin LTD, Vestar, Xerox, and Beckman Institute.

JA940273S

(18) Park, C.; Campbell, J. L.; Goddard, W. A., III. The monomer of the DNA binding region of the v-Jun leucine zipper protein recognizes the dimer binding site without dimerization. To be submitted for publication.

(19) Konig, P.; Richmond, T. J. *J. Mol. Biol.* **1993**, *233*, 139−154.

(20) Luisi, B. F.; Xu, W. X.; Otwinowski, Z., Freedman, L. P.; Yamamoto, K. R.; Sigler, P. B. *Nature* **1991**, *352*, 497−505.

(21) Cuenond, B.; Schepartz, A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 1154−1159.

(22) Weiss, M. A. *Biochemistry* **1990**, *29*, 8020−8024.

(23) Patel, L.; Abate, C.; Curran, T. *Nature (London)* **1990**, *347*, 572−575.

Can the Monomer of the Leucine Zipper Proteins Recognize the Dimer Binding Site without Dimerization?

Changmoon Park,[†,‡,§] Judy L. Campbell,[‡,⊥] and William A. Goddard, III*[,†,‡]

Contribution from the Materials and Process Simulation Center, Beckman Institute (139-74), Division of Chemistry and Chemical Engineering (CN 9056), and Division of Biology, California Institute of Technology, Pasadena, California 91125

Received February 27, 1995. Revised Manuscript Received October 12, 1995[⊗]

Abstract: It is generally believed that leucine zipper regulatory proteins for DNA transcription recognize their DNA binding sites as *dimers* preformed in solution (and that the monomers do not bind specifically to these sites). To test this idea, we synthesized the 31-residue peptide v-Jun-br, which contains *only* the DNA binding region of the v-Jun *monomer*. Footprinting assays show that v-Jun-br *monomers specifically protect the DNA binding site of v-Jun in almost identically the same way as dimers*. Thus, (i) the monomer recognizes the half-site of the dimer binding site and (ii) dimerization does not appreciably affect the bound conformation of each monomer. These results may have implications in the regulation of transcription by such proteins. Thus, two monomers of v-Jun might bind sequentially to the dimer binding site followed by dimerization of v-Jun while bound. This may allow binding at concentrations too low for dimerization in solution.

1. Introduction

The molecular mechanism by which cells adapt their phenotype in response to external stimuli is of great interest in modern biology. A crucial role in modulating gene expression is likely played by the products of proto-oncogenes, a number of which reside in the nucleus. Properties commonly exhibited by such nuclear oncogenes include (a) rapid (often transient) induction in respone to numerous agents, (b) messenger RNA with a short half-life, and (c) a short half-life for the proteins encoded by the nuclear oncogene.[1] Fos and Jun (both members of the leucine zipper protein family) have been observed as the products of immediate−early induced genes in response to external stimuli.[2−4]

Leucine zipper proteins bind to DNA as a dimer, and it is believed that the dimerization of leucine zipper protein is a prerequisite to specifically recognizing the binding sites.[5,6] However, the short lifetime of such nuclear oncogenes raises questions as to whether the concentrations are suitable for dimerization in solution.

We report herein evidence that the leucine zipper basic region of v-Jun can bind as monomers to the dimer binding site. We suggest that this may be the dominant process at low concentrations. Section 2 summarizes previous experiments and conclusions concerning the binding mechanism. Section 3 discusses details for the experiments reported herein, while section 4 reports the results. Section 5 covers kinetics issues relating to the mechanisms of binding, and section 6 contains further discussion.

2. DNA Binding Mechanism of Leucine Zipper Proteins

Leucine zipper proteins have about 60 residues with the C-terminus containing a leucine zipper region (4 or 5 leucines occurring every 7 residues) responsible for dimerization and the N-terminus containing a basic region (about 30 residues) responsible for DNA binding.[7,8] The leucine zipper proteins dimerize by using the leucine zipper region to form a coiled-coil structure for the dimer.[8,9] Most mutant leucine zipper proteins unable to carry out dimer formation fail to recognize the binding site.[10−12] Many leucine zipper proteins which have mutations on the basic region also fail to bind to the specific DNA site even though the mutants can form heterodimers with other wild-type leucine zipper monomers.[5,6] Therefore, it is believed that the dimerization of leucine zipper protein is a prerequisite to specific recognition of the binding sites. This idea is supported by the observation that the oxidized dimer of the GCN4 basic region specifically recognizes the GCN4 dimer binding site, but the monomer does not.[13,14]

While carrying out a project aimed at designing new long DNA binding proteins,[15−18] we observed that the monomer of

* To whom correspondence should be addressed.
† Materials and Process Simulation Center, Beckman Institute.
‡ Division of Chemistry and Chemical Engineering.
§ Current address: Department of Chemistry, Chungnam National University, Taejon, South Korea.
⊥ Division of Biology.
⊗ Abstract published in *Advance ACS Abstracts*, April 1, 1996.

(1) Ransone, L. J.; Verma, I. M. *Annu. Rev. Cell Biol.* **1990**, *6*, 539−557.
(2) Greenberg, M. E.; Ziff, E. B. *Nature* **1984**, *311*, 433−435.
(3) Lamph, W. W.; Dwarki, V. J.; Ofir, R.; Montminy, M.; Verma, I. M. *Nature* **1988**, *334*, 629−631.
(4) Ryder, K.; Lau, L. F.; Nathans, D. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 1487−1491.
(5) Gentz, R.; Rauscher, F. J.; Abate, C.; Curran, T. *Science* **1989**, *243*, 1695−1699.
(6) Ranson, L. J.; Visvader, J.; Wamsley, P.; Verma, I. M. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 3806−3810.
(7) Vinson, C. R.; Sigler, P. B.; McKnight, S. L. *Science* **1989**, *246*, 911−916.
(8) Ellenberg, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell* **1992**, *71*, 1223−1237.
(9) Rasmussen, R.; Benregnu, D.; O'shea, E. K.; Kim, P. S.; Alber, T. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 561−564.
(10) Neuberg, M.; Adamkiewicz, J.; Hunter, J. P.; Muller, R. *Nature (London)* **1989**, *341*, 243−245.
(11) Turner, R.; Tjian, R. *Science* **1989**, *243*, 1688−1694.
(12) Heeckeren, W. J.; Sellers, J. W.; Struhl, K. *Nucleic Acids Res.* **1992**, *20*, 3721−3724.
(13) Talanian, R. V.; McKnight, C. J.; Kim, P. S. *Science* **1990**, *249*, 769−771.
(14) Talanian, R. V.; McKnight, C. J.; Rufkovski, R.; Kim, P. S. *Biochemistry* **1992**, *31*, 6871−6875.
(15) Park, C.; Campbell, J. L.; Goddard, W. A., III. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9094−9096.
(16) Park, C.; Campbell, J. L.; Goddard, W. A., III *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 4892−4896.
(17) Park, C. Ph.D. Thesis, Chemistry, Caltech, May 1993.

S0002-7863(95)00653-6 CCC: $12.00

a Peptides

```
v-Jun-br:      S  QERIKAERKR  MRNRIAASKS  RKRKLERIAR
V-Jun-N : CGG  S  QERIKAERKR  MRNRIAASKS  RKRKLERIAR
v-Jun-C :      S  QERIKAERKR  MRNRIAASKS  RKRKLERIAR  GGC
```

b Oligonucleotides

```
o-CC: 5'-ctcagatccggatcctaggttaaacgATGAcgTCATcggtataggtcgagaattcggatcct-3'
      3'-gagtctaggcctaggatccaatttgcTACTgcAGTAgccatatccagctcttaagcctagga-5'

o-NN: 5'-ctcagatccggatcctaggttaaacgTCATcgATGAcggtataggtcgagaattcggatcct-3'
      3'-gagtctaggcctaggatccaatttgcAGTAgcTACGgccatatccagctcttaagcctagga-5'
```

Figure 1. Sequences of the protein (a) and oligonucleotides (b) used in the gel retardation and footprinting studies. The total length of each oligonucleotide is 62. Peptide v-Jun-br contains the basic region of v-Jun (amino acids 214−244).[30] Peptides v-Jun-br and v-Jun-C were prepared as described previously.[15−17] Peptide v-Jun-br was chemically synthesized and purified, and the purity was checked by mass spectroscopy at the Biopolymer Synthesis Center at the California Institute of Technology:[15,17] calculated, 3822.3; experimental, 3824.6.

the basic region of v-Jun binds selectively to the dimer binding site. These results, reported herein, suggest that under appropriate conditions (low concentrations) the dimerization of v-Jun proteins might occur by (i) first binding one monomer to the DNA binding site and then (ii) binding of the second monomer, followed by (iii) coupling of the leucine zippers of the bound monomers to form the bound dimer. If so, this mechanism might be particularly relevant for binding of short-lived DNA binding proteins.

Leucine zipper proteins dimerize via the leucine zipper regions, leading to a Y-shaped dimer where each arm is basic and recognizes half of the dimer DNA binding site. The basic region has no fixed conformation in solution, but changes into an α-helix when bound to the specific site.[19−22] This model has been confirmed by a recent X-ray crystal structure for the complex of DNA with GCN4 (another leucine zipper protein) homodimer[8] and for the complex of DNA with Jun/Fos heterodimer.[23] The X-ray studies show that the DNA binding site and the α-helix of the basic region of these leucine zipper proteins are both linear. However, depending on the nature of the binding site, other systems may bend.[24] In the gel electrophoresis using Jun heterodimer, a bent α-helix was proposed for the basic region of Jun to explain the DNA bending induced by the binding of Jun.[25]

Experiments using only the basic region of GCN4[13] or v-Jun[15−17] (without the leucine zipper region), but dimerized at the carboxy termini (denoted as pCC) by an added linker, showed that the basic region alone will recognize the dimer binding site (denoted o-CC). In addition, dimerization at the amino termini to form a rearranged protein (denoted pNN) leads to recognition of the rearranged oNN binding site.[15−17] These studies suggested that the α-helices are bent when bound to DNA.[15−17]

It is widely believed that protein dimerization is essential for leucine zipper proteins to effect specific DNA recognition. Evidence in favor of this view are the following observations: (i) Most mutations that prevent dimerization also prevent DNA binding.[10−15] (ii) A normal Jun and a mutant Fos on its basic region cannot recognize specific DNA sites even though they can make a heterodimer together.[5,6] (iii) GCN4 makes a stable dimer in the absence of the specific DNA binding site.[26] (iv) The oxidized dimer of the GCN4 basis region specifically recognizes the dimer binding site, but the reduced monomer does not.[13]

On the other hand, consider the following: (v) NMR experiments show that, in the absence of the specific DNA binding site, the lifetime of the GCN4 homodimer is between 10 ms and 1 s.[22] This shows that, in the absence of specific DNA, the GCN4 dimer is not stable in solution. (vi) Competition experiments show that peptides containing only the basic region of Jun, Fos, and CREB retain their promoter selectivity.[6,27] (vii) LexA binds to DNA as a dimer, but the monomer of LexA also recognizes the half-site of the full dimer binding site.[28] (viii) Skn-1 which contains a basic region similar to those of leucine zipper proteins, but lacks a leucine zipper dimerization region, binds to specific DNA sequences as a monomer.[29]

3. Materials and Experiments

3.1. Peptides and Oligonucleotide Synthesis. In order to obtain a direct test of whether predimerization is essential for the binding of leucine zipper protein, we synthesized a peptide, v-Jun-br (Figure 1a), containing only the basic region of v-Jun monomer and carried out footprinting assays for oligonucleotides containing the dimer binding site.

Peptide monomers v-Jun-br, v-Jun-N, and v-Jun-C were chemically synthesized and purified as described previously[15,17] (see the caption for Figure 1). The automated stepwise syntheses were done on an Applied Biosystems Model 430A peptide synthesizer with an optimized synthetic protocol for the *N-tert*-butoxycarbonyl (*t*-Boc) chemistry. The peptides were purified by reversed-phase high-performance liquid chromatography (HPLC) on a Vydac C18 column. A linear gradient of 0−50% aqueous/acetonitrile/0.1% trifloroacetic acid was run over 120 min.

The procedure to synthesize homodimer pCC (and pNN) is done in oxidizing conditions (5 mM oxidized dithiothretiol). v-Jun-C (or v-Jun-N) dimerizes to form pCC (or pNN) which was purified by HPLC.

The oligonucleotides o-CC and o-NN (Figure 1b) were synthesized using the facilities at the Biopolymer Synthesis Center at Caltech and purified as described.[16,17] o-CC has the binding site (ATGAcgTCAT) of the v-Jun dimer while o-NN has a rearranged half-site (TCATcg-ATGA; see Figure 1b). The synthesized oligonucleotides were purified using 10% denaturing polyacrylamide gel, and duplexes were made between complementary oligonucleotides.

3.2. Footprinting Assays. The footprinting assay solution (in 50 μL) contained bovine serum albumin at 100 mg/mL, 5% glycerol, 20

(18) Park, C.; Campbell, J. L.; Goddard, W. A., III. *J. Am. Chem. Soc.* **1995**, *117*, 6287−6291.
(19) Weiss, M. A. *Biochemistry* **1990**, *29*, 8020−8024.
(20) Patel, L.; Abate, C.; Curran, T. *Nature (London)* **1990**, *347*, 572−575.
(21) O'Neil, K. T.; Hoess, R. H.; DeGrado, W. F. *Science* **1990**, *249*, 774−778.
(22) Weiss, M. A.; Ellenberg, T.; Wobbe, C. R.; Lee, J. P.; Harrison, S. C.; Struhl, K. *Nature (London)* **1990**, *347*, 575−578.
(23) Glover, J. N. M.; Harrison, S. C. *Nature* **1995**, *373*, 257−261.
(24) Konig, P.; Richmond, T. J. *J. Mol. Biol.* **1993**, *233*, 139−154.
(25) Kerppola, T. K.; Curran, T. *Science* **1991**, *254*, 1210−1214.
(26) Hope, I. A.; Struhl, K. *EMBO J.* **1987**, *6*, 2781−2784.
(27) Busch, S. J.; Sasson-Corsi, P. *Oncogene* **1990**, *5*, 1549−1556.
(28) Kim, B.; Little, J. W. *Science* **1992**, *255*, 203−205.
(29) Blackwell, T. K.; Bowerman, B.; Priess, J. R.; Weintraub, H. *Science* **1994**, *266*, 621−628.

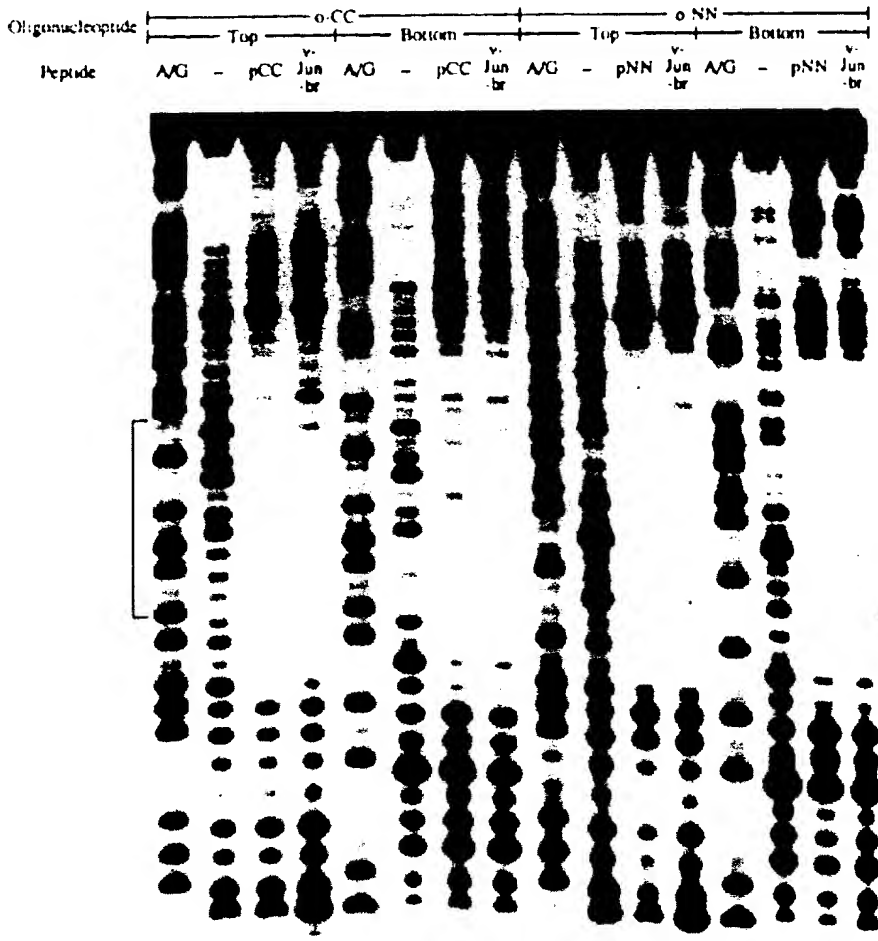

Figure 2. DNase I footprinting assays of v-Jun-br with oligonucleotides oCC and oNN. In order to compare the results of protection between monomer and dimer, DNase I footprinting assays of pCC and pNN were also carried out together with oCC and oNN, respectively. The brackets show the expected dimer binding sites (see Figure 1b). Peptide concentrations were determined as described previously.[15,17] A 50 000 cpm sample of each 5′-^{32}P-labeled probe DNA, bovine serum albumin (BSA) at 0.1%, poly(dI·dC) at 2 μg/mL, and 600 nM of pCC (or pNN) or 3 μM v-Jun-br (where indicated) were used in 50 μL of footprinting reaction solution as described previously.[15,17]

mM Tris-HCl (pH 7.5), 4 mM KCl, 2 mM $MgCl_2$, 1 mM $CaCl_2$, poly-(dI·dC) at 2 μg/mL, 50 000 cpm of each 5′-^{32}P-labeled probe DNA (about 20 fmol), and 0.6 μM pCC (or pNN) or 3.0 μM v-Jun-br where indicated. This solution was stored at 4 °C for 1 h. After adding 5 μL of DNase I diluted in 1× footprinting assay buffer, the solutions were stored for 1 min more at 4 °C. The DNase I digestion was stopped by addition of 100 μL of DNase I stop solution containing 15 mM EDTA (pH 8.0), 100 mM NaCl, 25 μg/mL sonicated salmon sperm DNA, and 25 μg/mL yeast tRNA. This was phenol/chloroform extracted, ethanol precipitated, and washed with 70% ethanol. The pallet was resuspended in 5 μL of formamide loading buffer, denatured at 90 °C for 4 min, and analyzed on 10% denaturing polyacrylamide sequencing gel (50% urea).

4. Results

The footprinting assays (Figure 2) show that the monomer v-Jun-br protects identically the same site as the dimer pCC (and pNN). (a) Columns 3 and 7 show that, for o-CC (top and bottom), the dimer pCC leads to recognition of the pCC binding site (marked with brackets). (b) Columns 4 and 8 show that monomer v-Jun-br also protects the complete pCC dimer binding site. (c) Columns 11 and 15 show that, for o-NN (top and bottom), the dimer pNN leads to recognition of the pNN binding site. (d) Columns 12 and 16 show that the monomer v-Jun-br also protects the complete pNN dimer binding site.

Because v-Jun-br contains only the basic region, there is no possibility of dimerization. Since the C-termini become positioned near each other when two monomers bind to the pCC binding site while the N-termini of both monomers are positioned near each other when two monomers bind to the pNN binding site, the similarity in the results between monomers and dimers shows that there are no specific interactions between the two monomers when bound to the site.

These results also indicate that the added linkers (Gly-Gly-Cys or Cys-Gly-Gly) when oxidized to form the dimer do not appreciably change the bound conformations of the monomers on the binding site of pCC (and pNN). Thus, each monomer retains the same contacts with DNA on both sites.[16,17]

These results also suggest that oxidization and covalent bonding of the thiol groups of the linkers to make the pCC and pNN dimers do not cause sufficient tension to change the contacts between the monomer and DNA.

5. Comparison between Dimer Formation in Solution and Dimer Formation on DNA

Figure 3 shows the relevant steps for two processes of forming bound DNA dimer: (a) Figure 3a considers that the dimer forms in solution, leading to an equilibrium constant of

$$K_D = k_{fMM}\, k_{rMM} = [D]\ ([M_1][M_2]) \qquad (1)$$

and the dimer binds to DNA, leading to an equilibrium constant of

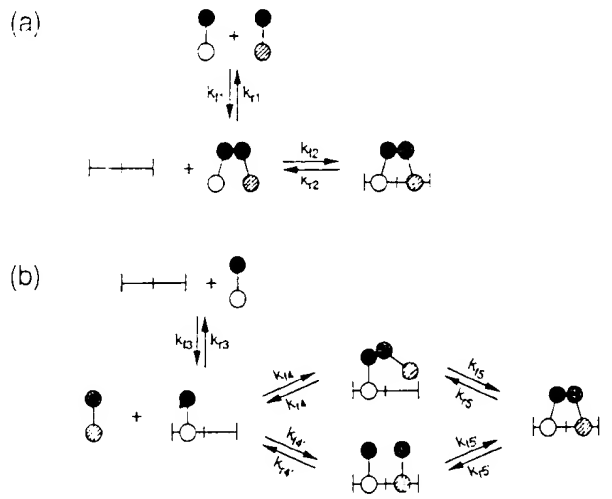

Figure 3. Two pathways for DNA binding of protein dimers: (a) dimer-only binding to the DNA binding site and (b) sequencial binding of two monomers to the DNA binding site. The darker (black and checked) circles represent dimerization regions and the brighter (white and striped) circles represent DNA binding regions (modeled after Figure 1 of Kim et al.[28]). k_f indicates the forward rate constant, and k_r indicates the reverse rate constant.

$$K_{DS} = k_{fDS}/k_{rDS} = [DS]/([D][S]) \quad (2)$$

(b) Figure 3b considers that the monomer binds to the dimer binding site, leading to an equilibrium constant of

$$K_{MS} = k_{fMS}/k_{rMS} = [M_1S]/([M_1][S]) \quad (3)$$

which is followed by binding of the second monomer. This second step may occur by two pathways (b1) dimerization of the leucine zipper of the free monomer to bound monomer, followed by binding of the second basic region to DNA, and (b2) binding of the second monomer to the dimer binding site, followed by dimerization of the leucine zipper regions.

In order to compare the two pathways a and b, consider the following kinetic scheme where M_i denotes the monomer, D denotes the dimer, and S denotes the DNA dimer binding site (the brackets indicate concentration). For pathway a we have

$$d[D]/dt = k_{fMM}[M_1][M_2] \quad (4)$$

$$d[D-S]_a/dt = k_{fDS}[S][D] \quad (5)$$

$$= K_D k_{fDS}[S][M_1][M_2] \quad (6)$$

where eq 1 was used. For pathway b1, we consider

$$d[M-S]/dt = k_{fMS}[S][M_1] \quad (7)$$

$$d[D-S]_b/dt = k_{fM-MS}[M_1S][M_2] \quad (8)$$

$$= K_{MS} k_{fM-MS}[S][M_1][M_2] \quad (9)$$

In each case the forward rate constant is much greater that the backward rate constant. For example,

$$K_D = k_{fMM}/k_{rMM} \approx 5 \times 10^4\ M^{-1} \quad (10)$$

based on the results of NMR experiments for GCN4.[22]

Thus, for conditions in which the concentration of a product is not too high compared to the concentration of the reactants, the backward reactions can be ignored in deriving equilibrium equations.

Equations 4−9 lead to the following relative rate constants:

$$\frac{(d[D]/dt)}{(d[D-S]_a/dt)} = \frac{k_{fMM}}{K_D k_{fDS}[S]} \quad (11a)$$

$$\approx 1/(K_D[S]) \quad (11b)$$

$$\frac{(d[D-S]_a/dt)}{(d[M-S]/dt)} = \frac{K_D k_{fDS}[M_2]}{k_{fMS}} \quad (12a)$$

$$\approx K_D[M_2] \quad (12b)$$

$$\frac{(d[D-S]_a/dt)}{(d[D-S]_b/dt)} = \frac{K_D k_{fDS}}{K_{MS} k_{fM-MS}} \quad (13a)$$

$$\approx K_D/K_{MS} \quad (13b)$$

$$\frac{(d[M-S]/dt)}{(d[D-S]_b/dt)} = \frac{k_{fMS}}{K_{MS} k_{fM-MS}[M_2]} \quad (14a)$$

$$\approx 1/(K_{MS}[M_2]) \quad (14b)$$

where eqs 11b, 12b, 13b, and 14b assume that the forward rate constants are similar (binding a monomer or a dimer to the DNA binding site).

If it is assumed that the dimerization rate constant of the monomers (k_{fMM}) is fast enough to provide dimers whenever they are needed, the binding of a dimer to the DNA binding site will be the rate-determining step in pathway a. From eq 14b, the rate-determining step for path b depends on the product of the concentration of monomer M_2 and the equilibrium constant of monomer binding to the DNA binding site. Thus eqs 7 and 9 becomes equal when $[M_2] = 1/K_{MS}$. From eq 13b, the relative rate constant for forming a complex between the dimer and the dimer binding site for path a to that for path b is equal to K_D/K_{MS}.

These equations allow an estimate to be made for the time to form the DNA bound dimer. At low concentration of monomers M_1 and M_2 ($<10^{-7}$ M), the DNA binding reaction for path a depends on the dimer binding reaction, while for path b the monomer binding to the monomer bound DNA binding site is the rate-determining step (assuming $K_{MS} \approx 10^6\ M^{-1}$ from ref 28).

At high concentration of monomers ($>10^{-5}$ M) path a (which involves formation of a dimer complex followed by binding of the complex to the dimer binding site) becomes faster than path b (from eq 12) because of the high population of protein dimers in solution. However, for a low concentration of monomers, the monomer binding mechanism (path b) leads to a net rate increase of 10−100 times [depending on the ratio of K_D and K_{MS} (see Figure 3b)] for forming a complex of two monomers at the DNA binding site compared to the dimer-only binding mechanism (path a). (In the case of LexA, a rate increase of about 75 times is proposed under their experimental conditions.[28]) Because the rate constant of binding the dimer complex to the dimer binding site depends on the concentration of both monomers (as in eqs 2 and 4), reaction through path b leads to a larger rate for complex formation when the concentration of

either monomer is very low (as in the case of Jun and Fos where heterodimers are made between them).

6. Discussion

It has been believed that leucine zipper proteins recognize their DNA binding sites as dimers which are preformed in solution and that monomers do not bind selectively to the DNA binding sites.[13,26] However, our current results (Figure 2) show that the monomer of the v-Jun basic region (v-Jun-br) specifically binds to both halves of both dimer binding sites o-CC and o-NN. Because v-Jun-br has no functional motif to become a dimer and because it recognizes both the o-CC and o-NN binding sites, we conclude that v-Jun-br recognizes the half-site of the dimer binding site as a monomer even though it has much weaker binding affinity to specific DNA sites compared to a dimer. These results are consistent with competition experiments which show that peptides including only the basic region of Jun, Fos, and CREB compete with the intrinsic Jun/Fos and CREB in DNA binding.[27]

These results contrast with the situation for GCN4 where only the dimer binds. This difference could be because v-Jun binds to DNA in a conformation different from that of GCN4.

Indeed residues on the carboxy terminus of the basic region of various leucine zipper proteins differ greatly from each other while the residues of the rest of the basic region are highly conserved.[7,30] Thus, mutations on the terminal residues of Fos substantially reduced the DNA binding affinity.[31] In contrast, the terminal residues of GCN4 do not show any direct involvement in DNA binding.[8,23] Therefore, the terminal residues may be responsible for the difference in behavior among leucine zipper proteins (as proposed by refs 18 and 32).

Experiment shows that the basic region of Jun competes with the Jun/Fos heterodimer in DNA binding.[6] This suggests that the Jun basic region recognizes the specific DNA site. Experimental results on the heterodimer formed between a wild-type Jun and a mutant Fos might seem inconsistent. This mutant Fos lacks the ability to bind to specific DNA sites but is still able to form a heterodimer with a Jun monomer that cannot recognize the specific DNA site.[5,6] This apparent discrepancy can be rationalized because the much weaker DNA binding affinity of a monomer as compared to a dimer might prevent detection of the monomer during gel retardation asseys at the concentrations used.

Our results[17] are consistent with a recent study[28] on the DNA binding protein LexA, which as a dimer recognizes a site having dyad symmetry. Kim et al.[29] showed that the standard dimer binding mechanism does not explain the fast binding rates of DNA binding proteins when equilibrium constants of dimerization of monomers are too low to provide appropriate concentrations of dimers in solution. Kim et al. proposed the mechanism in Figure 3b for the binding of LexA proteins to their DNA binding sites. In this proposed DNA binding mechanism, a monomer first binds to the binding site and dimerization with a second LexA occurs on the DNA binding site. The dissociation constant of LexA is similar to that of leucine zipper protein for both complex formation[28,33] between protein and DNA and protein dimerization.[22,34] Our results[17] are also consistent with experimental results[29] which show that the Skn-1 basic region binds to DNA as a monomer. The basic region of Skn-1 shows greater homology with Jun than ours to GCN4.

7. Summary

For both pCC and pNN binding sites, the monomer and dimer of v-Jun-br both lead to complete protection of the binding site with the same length of protected region. This suggests that v-Jun might dimerize on the binding site, removing the prerequisite of dimerization before binding. This could have profound implications in the regulatory mechanisms involving leucine zipper proteins. For example, it could allow binding at concentrations too low for dimerization in solution.

Acknowledgment. This research was supported by a grant from the Biological and Chemical Technology Research (BCTR) Program (David Boron) of the Department of Energy. The facilities of the Materials and Molecular Simulation Center (MSC) are also supported by grants from the National Science Foundation (CHE 94-13930 and ASC 92-17368), Asahi Chemical, Asahi Glass, Chevron Petroleum Technology Co., BF Goodrich, BP Chemical, Vestar, Hughes Research Laboratories, Xerox, and Beckman Institute.

JA950653T

(30) Struhl, K. *Cell* **1987**, *50*, 841−846.
(31) Neuberg, M.; Schuermann, M.; Muller, R. *Oncogene* **1991**, *6*, 1325−1333.
(32) Alber, T. *Curr. Biol.* **1993**, *3*, 182−184.
(33) Hope, I. A.; Struhl, K. *Cell* **1985**, *43*, 177−188.
(34) Schnarr, M.; Ponyet, J.; Granger-Schnarr, M.; Daune, M. *Biochemistry* **1985**, *24*, 2812−2818.